

Developing a Capability Approach to Measure and Value Quality of Life: An application to chronic pain

Philip B. Kinghorn

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia, Health Economics Group, School of Medicine,
Health Policy & Practice

June 2010

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the author's prior, written consent

Abstract

Resources used to produce healthcare are finite; the framework used to inform decisions regarding rationing is economic evaluation. In the UK, cost-utility analysis is recommended by NICE, but there are at least three problems with this approach stemming from the use of health as the evaluative space: (i) limited scope to compare health interventions with interventions falling under different policy contexts; (ii) focussing only on health is to omit other useful and important information; (iii) the maximization of health makes the separation of efficiency and equity impossible. The first of these is especially limiting given the current policy focus on 'joined up government'. Cost-benefit analysis theoretically offers the broadest evaluative space and allows consistency in evaluation across different contexts, but is hampered by methodological difficulties, especially within health.

The Capability Approach is explored as an alternative means of assessing outcomes for economic evaluation. Although the approach is attracting increasing attention in health economics, the theory is often described as incomplete. This thesis explores the steps involved with operationalising the approach in the context of chronic pain and discusses the problems and benefits encountered.

Qualitative work is conducted with chronic pain patients to identify important capabilities, which are incorporated into a questionnaire. Capability states are valued using a multi-attribute value method, a simple and choice-less approach, but one seemingly susceptible to part-whole bias. With an awareness of the susceptibility of approaches other than willingness-to-pay to part-whole bias this work demonstrates that the issue can be minimised through careful study design.

Similarity between capabilities identified here and elsewhere suggests it may be possible to develop one list of generic capabilities to be used in the evaluation of any public sector intervention in the UK. Such a list would make the Capability Approach a credible alternative to cost-utility and cost-benefit analysis.

For my parents

Acknowledgements

Funding was received from the Pain Management Clinic at the James Paget University NHS Hospital and from a UEA studentship. Thanks are owed to Dr William Notcutt, Miranda Mugford and Richard Smith for securing the funding and arranging the studentship.

Assistance from staff at the Pain Management Clinic was vital in order to recruit patients for the first two phases of the project, and the kind support and patience of the team there, especially Dr William Notcutt and Cheryl Phillips, is much appreciated. The project would not have been possible without the involvement of patients from the Pain Management Clinic who gave their time so generously and who were willing to share their views and experiences. Recruitment for the third phase of the project involved utilising a database maintained by the School of Economics at UEA and I would like to thank Judith Metha for allowing and facilitating access to this database.

Aspects of the project have been discussed at meetings of the Health Economists' Study Group and International Health Economics Association and useful suggestions and feedback received. Support from and the ability to share ideas with Jo Coast, Paula Lorgelly, Terry Flynn and Mozaffar Qizilbash has been particularly useful.

I am especially grateful to Angela Robinson and Richard Smith for supervising the thesis, for their guidance, support and encouragement. Finally, I am grateful to my examiners, Jo Coast and Richard Holland, for their feedback and suggestions.

Contents

Abstract.....	2
Contents.....	5
List of Tables.....	11
List of Figures.....	13
List of Abbreviations.....	15
CHAPTER 1: General Introduction.....	16
1.1 Economic Evaluation and the Policy Agenda in the UK.....	17
1.2 The Capability Approach: ‘A Third Way’.....	20
1.3 Chronic Pain.....	22
1.4 Health versus Quality of Life.....	24
1.5 Using the Capability Approach to Assess Quality of Life in Patients with Chronic Pain.....	26
1.6 Summary of Results and Contributions to the Literature.....	27
CHAPTER 2: Chronic Pain, Its Impact on Quality of Life and Assessment by Disease Specific Outcome Measures.....	30
2.1 The Effects of Chronic Pain on Quality of Life.....	31
2.2 Common Outcome Measures for Chronic Pain.....	37
2.3 Summary.....	45
CHAPTER 3: Measuring Outcome for Economic Evaluation.....	47
3.1 Introduction.....	47
3.2 Welfarist Approach.....	50
3.3 Non-Welfarism.....	57
3.3.1 Cost-Effectiveness Analysis.....	57
3.3.2 Cost-Utility Analysis.....	58
3.4 The Importance of the Evaluative Space.....	63

3.4.1 Health.....	63
3.4.2. Money as a Proxy for Utility.....	68
3.4.3 Benefits Explicitly Valued in WTP Studies.....	71
3.4.4 Experienced Utility.....	72
3.5 Perspective.....	73
3.5.1 Cost-Utility Analysis.....	73
3.5.2 Cost-Benefit Analysis.....	75
3.6 Discussion.....	76

CHAPTER 4: Systematic Review: Outcome Measures Used in Economic Evaluations Relating to Chronic Pain.....78

4.1 Introduction.....	78
4.2 Search Methods.....	80
4.3 Search Results.....	85
4.4 Discussion.....	93
4.4.1 EuroQol EQ-5D.....	96
4.4.2 Short Form-36 and Short Form-12 Health Survey.....	99
4.4.3 Other Outcome Measures.....	101
4.4.4 Studies in Which Parameters were Identified through Searches of the Literature.....	102
4.5 Summary.....	105
4.6 Conclusion.....	108

CHAPTER 5: The Capability Approach.....111

5.1 The Capability Framework.....	112
5.1.1 Specifying the evaluative space.....	113
5.1.2 Happiness & desire fulfilment.....	115
5.1.3 Primary goods.....	116
5.2 Operationalising the Capability Approach.....	117
5.2.1 Functionings or Capabilities?.....	118
5.2.2 The Selection of Relevant Capabilities.....	120
5.3 Intrapersonal Aggregation.....	124
5.3.1 Is weighting necessary?.....	124
5.3.2 Equal Weights?.....	125

5.4 The Practical and Conceptual Complexity of the Weighting Task.....	127
5.4.1 Happiness and desire fulfilment.....	127
5.4.2 Utility as choice.....	129
5.4.3 Scope for a little pragmatism.....	129
5.5 Capability and Health.....	131
5.6 Summary.....	135

CHAPTER 6: Systematic Review: Practical Applications of the Capability Approach.....137

6.1 Introduction.....	137
6.2 Search Methods.....	139
6.3 Search Results.....	142
6.4 Discussion.....	147
6.4.1 The Different Contexts in which the Capability Approach has been Applied.....	148
6.4.2 Obtaining Lists of Capabilities.....	150
6.4.3 Judging Achievement and Calculating an Index of Capability.....	162
6.5 Summary.....	168

CHAPTER 7: Exploring Possible Methods for Weighting Capabilities..171

7.1 Requirements of a Valuation Method.....	172
7.2 Willingness to Pay.....	176
7.3 Discrete Choice Experiments.....	177
7.4 Multi-Attribute Utility Methods.....	178
7.5 Adopting the MAU Approach in the Context of Capability.....	180
7.6 Choosing a Functional Form for the MAV Model.....	182
7.7 Summary.....	184

CHAPTER 8: Aims and Objectives of the Empirical Work.....186

8.1 Motivation.....	186
8.2 Aims & Objectives.....	187
8.2 Identifying Capabilities of value to Chronic Pain Patients.....	189
8.3 Assessing Achievement in terms of Capability.....	189
8.4 Determining the Value of Capabilities.....	190

8.5 Summary.....	190
------------------	-----

CHAPTER 9: Study Phase I: Identifying Capabilities and Developing a Capability Instrument.....191

9.1 Selecting a Means of Collecting Qualitative Data.....	191
9.2 Focus Group Design.....	192
9.3 Initial Recruitment.....	195
9.3.1 Participant Characteristics.....	196
9.4 Framework Analysis.....	197
9.4.1 Data Management.....	198
9.4.2 Summarising & Assigning Meaning.....	200
9.4.3 Refining Categories & Classifying Data.....	205
9.5 Individual Interviews.....	221
9.5.1 Feedback from the Interviews.....	223
9.5.2 Similarities with Lists of Capabilities Developed Elsewhere.....	227
9.6 Questionnaire Design.....	229
9.6.1 Other Material Contained in the Questionnaire.....	230
9.7 Pilot Stage.....	230
9.8 Summary.....	236

CHAPTER 10: Study Phase II: Trialling a Capability Instrument.....239

10.1 Recruitment.....	240
10.2 Follow-Up Questionnaires.....	241
10.3 Participant Characteristics.....	242
10.4 Results.....	243
10.4.1 Item Response Rate.....	243
10.4.2 Exploring Respondents' Understanding of the Questions.....	247
10.4.3 Comparability between the Two Instruments.....	249
10.5 Discussion.....	252

CHAPTER 11: Study Phase III: Weighting Capabilities.....257

11.1 Study Overview.....	258
11.2 Sample Size.....	259

11.3 Recruitment.....	260
11.4 Scaling Exercise for Levels.....	262
11.5 Swing-Weighting Exercise.....	264
11.6 Study Design & Statistical Methods.....	268
11.7 Results: Scaling Exercise.....	273
11.8 Results: Swing-Weighting Exercise.....	277
11.8.1 Mean Response Method of Calculating Importance Weights for Capabilities.....	287
11.8.2 Using Regression Analysis to Derive Importance Weights.....	289
11.9 Calculating a Capability score for Respondents.....	291
11.10 Part-Whole Bias.....	297
11.10.1 Amendments to the Study Design.....	299
11.10.2 Results: Study Extension.....	301
11.11 Summary.....	303
CHAPTER 12: Discussion & Further Recommendations.....	306
12.1 Reflections on the Qualitative Work.....	307
12.2 Theoretical Issues & Contributions.....	311
12.2.1 Health Versus Quality of Life.....	311
12.2.2 Lists of Capabilities.....	311
12.2.3 Broader Implications of the Research.....	314
12.3 Policy Implications.....	315
12.4 Conclusion.....	317
References.....	319
Appendices.....	A1
Appendix A.....	A2
Invitation Letter – Focus Groups (Study Phase I).....	A3
Information Sheet – Focus Groups (Study Phase I).....	A4
Consent Form – Focus Groups (Study Phase I).....	A8
Preliminary Information Sheet (Study Phase I).....	A10
Topic Guide for Focus Groups (Study Phase I).....	A12

Appendix B.....	A14
Invitation Letter – Piloting of Questionnaire (Study Phase II).....	A15
Information Sheet – Piloting of Questionnaire (Study Phase II).....	A16
Consent Form – Piloting of Questionnaire (Study Phase II).....	A20
Feedback Form (Study Phase II).....	A22
Appendix C.....	A24
Initial Capability Questionnaire.....	A25
Follow-up Capability Questionnaire.....	A46
Reminder Postcard.....	A66
Appendix D.....	A67
Invitation Letter (Study Phase III).....	A68
Information Sheet (Study Phase III).....	A69
Consent Form (Study Phase III).....	A72
Example Test and Questions from Participant Pack (Study Phase III).....	A73
Appendix E.....	A89
Links between Capability Questionnaires Developed within the Context of Health.....	A89

List of Tables

Table 2.1	Common Outcome Measures for Chronic Pain.....	37
Table 2.2	Inclusion of Dimensions Important to Chronic Pain Patients.....	44
Table 3.1	Measurement of Outcomes in Economic Evaluation.....	49
Table 4.1	Exclusion Criteria.....	82
Table 4.2	Inclusion Criteria.....	83
Table 4.3	Search Terms and Summary of Search Results: Economic Evaluation.....	84
Table 4.4	Economic Evaluations Relating to Chronic Pain.....	87
Table 4.5	Study Characteristics.....	91
Table 6.1	Exclusion Criteria: Capability Approach.....	140
Table 6.2	Inclusion Criteria: Capability Approach.....	140
Table 6.3	Search Terms and Summary of Search Results: Capability.....	141
Table 6.4	A Summary of Practical Applications of the Capability Approach.....	144
Table 6.5	Some Examples of Commodities/Activities, their Characteristics and the Achievements they help to Facilitate.....	157
Table 6.6	Nussbaum's Capabilities and Selected British Household Panel Survey Questions.....	160
Table 6.7	Criteria for Scoring Women's Participation.....	164
Table 6.8	Components of a Composite Measure of Deprivation.....	166
Table 9.1	Focus Group Participants.....	197
Table 9.2	Examples of Sub-Themes with References from Segment Four.....	200
Table 9.3	A Ranking of the Ten Capabilities (highest to lowest).....	225
Table 9.4	Ranking of the Ten Capabilities Based on EQ-VAS Scores.....	225
Table 9.5	Summary of Respondent Characteristics.....	233
Table 9.6	Respondent Characteristics.....	233
Table 9.7	Amendments made to Questionnaire.....	236
Table 10.1	Age & Income.....	242
Table 10.2	Employment.....	243
Table 10.3	Number of Incomplete Responses.....	244
Table 10.4	Number of Entries Missing: EQ-5D.....	244
Table 10.5	Number of Entries Missing: Initial Capability Questionnaire.....	245
Table 10.6	Number of Entries Missing: Follow-Up Capability Questionnaire.....	246
Table 10.7	Summary of the Four Groups Defined by Tariff Values at the Initial Stage.....	249
Table 10.8	Responses from the Capability Questionnaire by Group at Initial Stage.....	251
Table 10.9	Ability to Do Things Considered to be Worthwhile & Productive.....	252
Table 10.10	Social Interaction.....	252
Table 11.1	Summary of Participant Characteristics by Group.....	261
Table 11.2	Summary of Participant Characteristics.....	261
Table 11.3	Possible Scenarios if the Level of Only Three Capabilities is Varied.....	271
Table 11.4	Scenarios for Pilot of Swing Weighting.....	272
Table 11.5	Scores for Levels within Dimensions.....	274
Table 11.6	Summary Statistics for Table 10.5.....	276
Table 11.7	Bottom-Up Scores with Only One Capability at Best.....	277
Table 11.8	Individual Dominance Tests.....	283
Table 11.9	Relationship between Number of Categories at Best and Score.....	283
Table 11.10	Mean Scores by Group.....	285
Table 11.11	Mean Scores with Cleaned Data.....	285
Table 11.12	Weights for Selected Capabilities by Mean Response Method.....	287

Table 11.13	Comparison of Importance Weights for Capabilities by Mean Response and by Regression Analysis.....	290
Table 11.14	Comparison of Rank Ordering of Capabilities via Three Different Methods.....	291
Table 11.15	Importance Weights.....	293
Table 11.16	Weighted Values for Levels within Functionings (D_{WZ}).....	295
Table 11.17	Characteristics of Participants (Extension: Study Phase Three).....	301
Table 11.18	Summary of Characteristics of Participants (Extension: Study Phase Three).....	301
Table 11.19	Relationship between Number of Capabilities at Best and Score Attached to Scenario (Extension: Study Phase Three).....	302
Table 11.20	Impact of Changes to Study Design.....	302
Table 11.21	Top-Down Scores with Only One Capability at Worst: Comparison.....	303

List of Figures

Figure 1.1	Pathway Through Which the Impact of Ill Health is Experienced.....	25
Figure 2.1	Questions from the Nottingham Health Profile.....	40
Figure 2.2	SF-6D.....	42
Figure 4.1	Summary of Literature Search for Articles Discussing/Detailing Economic Evaluations for the Relief of Chronic Pain.....	85
Figure 4.2	Parameters Included in the Model by Neighbors <i>et al.</i>	103
Figure 5.1	Commodities, Capabilities and Functionings.....	113
Figure 5.2	Value Purposes.....	114
Figure 6.1	Summary of Literature Search: Capability.....	142
Figure 6.2	New Capabilities in the Christchurch Case Study.....	152
Figure 6.3	New Capabilities in the Samoa Case Study.....	152
Figure 6.4	The Sequence from Commodities to Utility.....	158
Figure 7.1	Weighting Attributes: Bottom-Up Method.....	182
Figure 9.1	The Stages and Processes Involved in Framework Analysis.....	198
Figure 9.2	Main Themes (Initial).....	199
Figure 9.3	Refining List of Main Themes.....	201
Figure 9.4	An Initial Summary of the Main Theme 'Employment'.....	202
Figure 9.5	An Initial Summary of the Main Theme 'Independence & Control'.....	203
Figure 9.6	Initial List of Ten Higher-Order Capabilities and Associated Functionings.....	208
Figure 9.7	A Summary and Illustration of the Capability: To be able to achieve self-respect.....	210
Figure 9.8	A Summary and Illustration of the Capability: To have reasonable opportunity to enjoy social interaction.....	212
Figure 9.9	A Summary and Illustration of the Capability: To be able to fulfil the role of parent/grandparent.....	213
Figure 9.10	A Summary and Illustration of the Capability: To have the ability to spend time with and feel well supported by your family.....	214
Figure 9.11	A Summary and Illustration of the Capability: To be able to remain both physically and mentally active.....	215
Figure 9.12	A Summary and Illustration of the Capability: To be recognised as a person with an individual identity – not be seen as an illness.....	216
Figure 9.13	A summary and Illustration of the Capability: To be able to feel independent and to feel like you have control over your own life.....	217
Figure 9.14	A Summary and Illustration of the Capability: Not be constrained in your ability to experience a loving relationship.....	218
Figure 9.15	A Summary and Illustration of the Capability: To be able to achieve a good level of physical and mental well-being.....	219
Figure 9.16	A Summary and Illustration of the Capability: To be able to find enjoyment in life.....	220
Figure 9.17	Final List of Higher-Order Capabilities.....	226
Figure 9.18	Questions Contained on the Feedback Form for Pilot Interviews.....	232
Figure 10.1	Histogram Illustrating the Overall Distribution of EuroQol Tariff Values.....	250
Figure 11.1	Example Question.....	262

Figure 11.2	Calculation in Gain in Value Associated with an Improvement in the Second Attribute.....	269
Figure 11.3	Mean Gains Obtained Using Bottom-Up.....	279
Figure 11.4	Mean Losses Obtained Using Top-Down.....	282
Figure 11.5	Comparison of Importance Weights for Capabilities by Mean Response and by Regression Analysis.....	290

List of Abbreviations

BDI	Beck Depression Inventory
BHPS	British Household Panel Survey
BPI	Brief Pain Inventory
BU	Bottom-Up Approach on Swing-Weighting Method
CA	Capability Approach
CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
CMA	Cost-Minimisation Analysis
CPG	Chronic Pain Grade
CUA	Cost-Utility Analysis
DCE	Discrete Choice Experimentation
EQ VAS	EuroQol Visual Analogue Scale
EU	Expected Utility
GNP	Gross National Product
HADS	Hospital Anxiety and Depression Scale
HDI	Human Development Index
HRQL	Health-related quality of life
HUI	Health Utility Index
MAUT	Multi-Attribute Utility Theory
NICE	National Institute of Health and Clinical Excellence
ODI	Oswestry Disability Index
OECD	Organisation for Economic Cooperation and Development
PQLI	Physical Quality of Life Index
QALY	Quality Adjusted Life Year
QWB	Quality of Well-Being
SF-6D	Short Form 6D
SG	Standard Gamble
TD	Top-Down Approach on Swing Weighting Method
TTO	Time Trade-Off
UNDP	United Nations Development Programme
VAS	Visual Analogue Scale
WTP	Willingness To Pay

CHAPTER 1: General Introduction

The resources used to produce health care are finite and devoting more resources to the treatment of one particular patient group results in resources being diverted away from some other use. No health care system, anywhere in the world, has achieved levels of spending sufficient to meet all its clients wants for health care (Morris, Devlin et al. 2007). The need for some form of rationing has been recognised by the UK Government and there is a commitment to base choices on robust evidence, with the aim of offering patients the best possible treatments that the health budget allows (House of Commons Health Committee 2008). The National Institute for Health and Clinical Excellence (NICE) is the cornerstone of this system. NICE is a quasi-independent organisation which assumes the role of decision maker, considering evidence and providing national guidance on the use of treatments, medicines and procedures (NICE 2009).

The economic framework used by NICE, and more widely within health economics, is economic evaluation; broadly “the comparative analysis of alternative courses of action in terms of both their costs and consequences” (Drummond, Sculpher et al. 2005, p9). Economic evaluation is an umbrella term, incorporating a range of specific approaches which largely differ only in the way in which benefits – or consequences – are defined and measured for inclusion in the analysis. It may not be obvious, in any given context, which of these approaches will be most appropriate. For example, NICE guidance for economic evaluation (NICE 2008) stresses the use of cost-utility analysis, where benefits are expressed in terms of Quality Adjusted Life Years (QALYs). In contrast, the Treasury’s ‘Green Book’ recommends cost-benefit analysis (HM Treasury). A single approach would deliver greater consistency across government departments, but while there are methodological difficulties associated with cost-benefit analysis, particularly when it is applied in a health care setting, there are also characteristics of cost-utility analysis which make it unsuitable for use outside a health care setting. This is important as the relationship between government departments within the

broad area of health gets closer, and the emphasis on ‘joined up government’ increases. This has, in part, led to renewed interest in other options open to the economist for benefit measurement in economic evaluation, such as cost-consequence analysis (Coast 2004; Coast, Smith et al. 2007), subjective well-being (experienced utility) (Dolan and Peasgood 2008) and Capability (Coast, Smith et al. 2008a; Grewal, Lewis et al. 2006; Lorgelly, Lorimer et al. 2008).

This thesis contributes to the newly emerging, but expanding, body of literature (Anand, Hunter et al. 2005; Anand and Van Hees 2006; Coast, Flynn et al. 2008; Kinghorn and Smith 2008; Coast, Smith et al. 2008a; Greco, Skordis et al. 2009; Kinghorn, Robinson et al. 2009) concerned with the application of the Capability Approach within health economics. This thesis aims to explore conceptual issues surrounding the Capability Approach and the steps which need to be taken in order to operationalise it. Methodologies suitable for implementing each step will be suggested and the appropriateness and practicality of these methodologies explored. The work will ultimately provide some indication as to the usefulness of the framework within health economics and given the findings recommendations will be made regarding further research. The development of the Capability Approach here is grounded in the assessment of the quality of life of patients with chronic pain.

1.1 Economic Evaluation and the Policy Agenda in the UK

Cost-benefit analysis can theoretically incorporate any cost and benefit and would allow policy interventions from any government department to be compared, using as it does the single and generic metric of money (Olsen and Smith 2001). Treasury guidelines therefore place considerable emphasis on the use of cost-benefit analysis in the decision-making process (HM Treasury). Where real market prices are unavailable, willingness-to-pay is the preferred approach. One report, by the Lords Economic Affairs Committee, states that: “cost-benefit analysis provides a useful framework ... and we endorse the idea that it should occupy a central place in risk assessment procedures” (Lords Economic Affairs Committee 2006, par. 95). Within the

same report it is acknowledged that due to “questions concerning the empirical and ethical robustness of the [cost-benefit] approach ... some public sector agencies in the UK have avoided their use” (Ibid. par. 56). One such agency is NICE, who instead advocate cost-utility analysis with the QALY. A QALY is calculated by multiplying length of life by a weight reflecting health-related quality life (HRQL). One QALY represents one year in perfect health, zero is used to represent death, and there can be negative values representing states worse than dead.

The NICE guidance for economic evaluation states (NICE 2008, p38):

“...the measurement of changes in HRQL should be reported directly from patients and the value of changes in patients’ HRQL (that is, utilities) should be based on public preferences using a choice-based method. The EQ-5D is the preferred measure of HRQL...”

The EQ-5D is a simple generic instrument, originally designed by the EuroQol Group to be used alongside other quality of life measures (Brooks 1996), but which is now often used as the sole measure of health-related quality of life, on which QALYs are based¹. The EQ-5D contains five dimensions: mobility; self-care; usual activities; anxiety and depression; pain and discomfort. Scores are generated for each through the use of three levels (normal functioning all, some or none of the time). A scoring algorithm then combines functioning across the five dimensions into a single weight to be used with life years spent in that state to generate the final QALYs (Brazier, Ratcliffe et al. 2007).

In a report by the House of Commons Health Committee (2008) a number of weaknesses with the QALY approach are acknowledged, namely the ambiguity about assumptions regarding what exactly constitutes the ‘quality’ for which life years are adjusted, and that QALYs may lead to bias against

¹ There are a variety of other instruments which can be used to generate QALYs (such as the SF-6D and the HUI), but the EQ-5D is the one recommended by NICE and the most popular in the UK, as well as being widely used in other countries. It is therefore the focus of the discussion relating to QALYs here.

treatments for chronic (long-term) conditions. Chapter three of this thesis will explore in detail cost-utility analysis with QALYs and it is found that far from weighting years of life according to some broad concept of quality of life, the evaluative space is commonly restricted to a narrow concept of health. This narrow evaluative space leads to at least three weaknesses: (i) limited scope to compare health interventions with interventions falling under different policy areas; (ii) to focus on health alone is to omit other useful and important information; (iii) the maximisation of health makes the separation of efficiency and equity impossible. Note that further explanation will be given in chapter three of how efficiency and equity are differentiated within a welfarist approach, but suffice to note at this point that in terms of health maximisation, the production of health cannot be separated from its allocation, but rather allocation happens simultaneously with production (Coast 2009).

The first of the three weaknesses listed above is highlighted in the Economic Affairs Committee report cited earlier (Lords Economic Affairs Committee 2006, par. 49):

[T]here appears to be considerable variation in methodology and practice between government departments. More attention therefore needs to be paid to ensuring greater consistency across departments and areas of application.

Consistency is particularly important given the overlap between health and social care services. From a patient perspective, the distinction between 'health needs' and 'social service needs' is often unclear (Grewal, Lewis et al. 2006). From a policy perspective, there has been much talk in various government White and Green papers of health, housing and social services working together to provide 'joined up care', and of local health and social care commissioners working together (HM Government 2006; HM Government 2009).

Cost-benefit analysis (willingness-to-pay) may then appear to be the most appropriate approach to use, avoiding as it does the narrow evaluative space and allowing consistency in the evaluation of projects conducted by any

government department. It will be seen in chapter three, however, that there are serious methodological issues with willingness-to-pay, such as part-whole bias and the effects of budget constraints. Within the context of health, the approach also tends to attract a high proportion of protest votes and patients may not always be well informed consumers, instead relying on clinicians to inform them about treatment options (Culyer 1991). This may explain the fact that no examples of cost-benefit analysis were found when a systematic review was conducted to find economic evaluations of treatments for chronic pain. Despite the limitations of the EQ-5D highlighted in chapters two and three, the instrument was used in a quarter of the economic evaluations identified for chapter four. In only one paper did the authors explain their reason for using the EQ-5D. It would appear from this review that the popularity of the instrument stems at least as much from the ready availability of utility values as it does from confidence in the descriptive system.

1.2 The Capability Approach: A 'third way'

It has been noted that there are methodological challenges associated with cost-benefit analysis and at least three weaknesses with cost-utility analysis. While researchers from both camps continue to address and attempt to overcome these challenges, it has been noted earlier in this chapter that there are other options available on which to base economic evaluation. This thesis will therefore explore and operationalise one such alternative means of assessing outcomes. This 'third way' is the Capability Approach, originally proposed and largely developed by Amartya Sen (Sen 1983; Sen 1987; Nussbaum and Sen 1993; Sen 1999a). The Capability Approach has become recognised in the development economics literature as an alternative to mainstream cost-benefit analysis as well as income based measures of well-being (Robeyns 2006). The approach is perhaps most well known for providing the theoretical grounding for the United Nations' Human Development Index (United Nations Human Development Programme 2006). It is also an approach attracting an increasing amount of attention in health economics (Verkerk, Busschbach et al. 2001; Cookson 2005; Coast, Flynn et al. 2006; Grewal, Lewis et al. 2006; Lorgelly, Lorimer et al. 2008; Coast, Smith

et al. 2008a). The strength of the approach, which focuses on the ability of individuals to lead a life that they have reason to value, is its wide informational base compared to rival approaches, which reduce well-being to a single category (utility, health or income) (Nussbaum and Sen 1993).

The Capability Approach relies on an important distinction between functionings and capabilities. Functionings are described by Sen as the various things that a person manages to do or be in leading a life (Nussbaum and Sen 1993, p31). The capability of a person is the alternative combinations of functionings the person can achieve, and from which he or she can choose one collection (Ibid.).

Although Sen has defined and refined his Capability Approach, in chapter five it will be seen that the theory is still often described as 'incomplete'. It has been noted and will be apparent from chapter six when practical applications of the approach are reviewed that the Capability Approach in practice comes in a variety of forms. In part this is because of the wide scope of the approach, but also because the approach is "radically underspecified" (Robeyns 2006, p353).

Sen suggests that the steps involved with operationalising the approach should involve the researcher considering a number of questions (Nussbaum and Sen 1993). The first of these is whether it is appropriate to work in terms of functionings or capabilities. Second, which functionings and capabilities are most important in any given context. Those which are most important should be selected to form the evaluative space. The final question concerns the relative importance of these functionings and capabilities.

In reality the incomplete nature of the approach leads to many more questions. For example, how should functionings and capabilities be selected (Alkire 2002; Robeyns 2006), who should assess a person's capability, who should decide the relative importance of different capabilities, and how (Sugden 1993)? In addition, there is the question of whether any instrument or interview will be able to accurately and realistically capture the underlying

ability to function, as opposed to just recording achieved functioning (Coast, Smith et al. 2008a). If the ability to achieve something is valuable then why would we not observe it as an outcome? And, likewise, if a person does not choose to perform something, how do we know if they are able to?²

For the Capability Approach to prove useful – or otherwise – it must first be used and some attempt made to apply it in practice in a health context. Only then will the issues be tackled and either solved or found to be of inescapable hindrance. This thesis will therefore explore what steps are necessary in operationalising the Capability Approach, find ways of implementing these steps, explore problems, issues and benefits along the way, and review the final results. This is especially important to undertake in light of the simultaneous initiatives being explored by researchers based at Birmingham and Glasgow, which have both taken different approaches to that outlined in this thesis. Triangulating these three experiences is a critical component of the discussion of the results of this thesis.

Given the very exploratory nature of the exercise, methods for implementing each step have been chosen from a range of possibilities, where there has been no clear right or wrong option. The quantitative results are not intended to be useful for inclusion in an economic evaluation, but the work is first and foremost intended to be exploratory. Lessons learnt may one day lead to results which *can* be considered alongside and combined with cost to inform policy decisions. However, it was important to ground the work in a practical context, and one which differed from simultaneous work elsewhere, and thus chronic pain was chosen.

1.3 Chronic Pain

The policy documents discussed earlier in this chapter tend to focus on the provision of joint health and social care services for three groups: children, older people and people with long-term conditions. A separate but related

² This point was raised by John Brazier at the International Health Economics Association Conference in Copenhagen, 2007.

stream of research has been conducted to identify and value capabilities relevant to older people (Grewal, Lewis et al. 2006; Coast, Flynn et al. 2008). Research is also underway to develop and refine a list of capabilities for use in public health contexts (Lorgelly, Lorimer et al. 2008). The Capability Approach has been used in this thesis as a framework within which to assess the quality of life of patients with chronic pain.

Chronic pain is defined as pain lasting for over six months and serving no useful purpose, unlike acute pain which acts as a warning that damage is being done to the body (Coniam and Diamond 1994; Gamlin and Lovel 2002). Acute pain prompts us to remove ourselves from danger, tells us that we have injured ourselves, and aids diagnosis in the case of disease. In most cases the pain will go when the cause is cured and the pain is no longer of any significance. All other pains generate suffering, cause anxiety and then depression, interfere with the ability to live a normal life and cause invalidity (Coniam and Diamond 1994). According to one UK forum, chronic pain affects 7.8million people in the UK of all ages (CPPC).

A number of factors set chronic pain aside as being of particular interest. As chronic pain is a long-term condition, whose cause is frequently not fully understood (Coniam and Diamond 1994), treatments in many cases involve masking the symptom of pain rather than curing the underlying cause. Treatment is therefore often protracted rather than taking the form of a one-off intervention. There is also vast variation in the types of pain people experience, the pain location and the degree of severity, and there are a wide range of treatments, from physiotherapy to epidurals, simple paracetamol to opioids and new cannabinoid based products. Finally, and possibly most importantly, chronic pain can affect virtually every aspect of a patient's daily life.

Given that cost-utility analysis and QALYs (especially those based on the EQ-5D) tend to have a narrow focus on health, it is likely that they may not fully capture everything that is important to patients with chronic pain and reflect the full impact on broader quality of life. This will be investigated in chapter

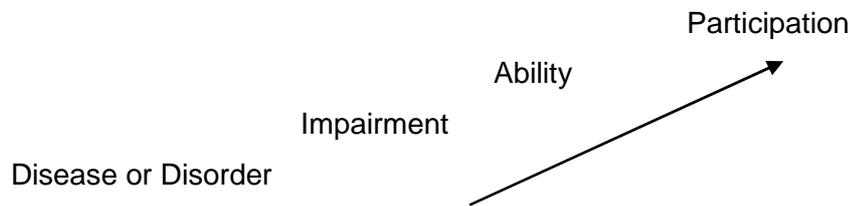
two, following a more comprehensive introduction to chronic pain. As chronic pain is a long-term condition, often causing invalidity, then patients are more likely to need support from social and housing services as well as the health system. Given that a cure is often not available then support from a range of agencies can help the patient to live more comfortably with their illness. Recall that the government believes these services can be delivered most effectively when providers coordinate their efforts, but with cost-utility analysis it is difficult to evaluate health interventions alongside those offered by other providers.

1.4 Health versus Quality of Life

Coast, Smith *et al.* (2007a) and Grewal, Lewis *et al.* (2006) note that economists' attempts to assist resource allocation decisions are currently focussed very strongly upon the measurement of health, with the use of proxies for health (life expectancy and health-related quality of life) dominating the empirical economic evaluation literature. It is noted in chapter three that many such measures of quality of life (not just those used in the formulation of QALYs) in fact measure one influence on quality of life (health) rather than quality of life more broadly. To some extent this is no surprise, health is the defining goal for a health system. But, it has already been stated within this chapter that the use of health as the evaluative space (or maximand) in cost-utility analysis makes it difficult or impossible to compare health interventions with interventions falling under different policy areas and that important outcomes may be excluded from the evaluation. Of course, the suitability of health as the maximand depends to some extent on how health is defined and what health is used as a proxy for.

The World Health Organisation (1948) has defined health as: 'A state of complete, physical, mental and social well-being, and not merely the absence of disease and infirmity'. More recently the WHO have defined health in terms of impairment, disability and handicap, and have set out a spectrum, or pathway, through which health is seen as having an impact; as illustrated in Figure 1 (Brazier, Ratcliffe *et al.* 2007).

Figure 1.1: Pathway Through Which the Impact of Ill Health is Experienced



Instruments designed to measure health will capture some concept or form of health that fits somewhere along the above spectrum. For example, we could assess pain severity, the impact of this pain on mobility, or the impact this has on a patient's involvement in society. While some instruments focus on impairment, others such as the EQ-5D contain dimensions which map to different points on the spectrum. For example, the EQ-5D contains two dimensions relating to impairment (pain and depression and anxiety), two relating to ability (mobility and self-care), and one relating to participation (usual activities) (Brazier, Ratcliffe et al. 2007).

Contrary to this classification by Brazier, Ratcliffe *et al.* usual activities could be considered to relate to multiple levels along the pathway, namely impairment, ability and participation. Whereas four of the dimensions are clearly and narrowly defined, 'usual activities' appears to be relied upon as a catch-all term to capture anything else important to the respondent. There is no direction as to which aspects of life respondents should interpret as being usual activities and no reason to believe that the interpretation across respondents will be consistent.

In terms of quality of life, we are not helped by the lack of any consensus on what the definitions of either quality of life or well-being are and what they include (Kind 1990; Megone 1990; Dolan and Peasgood 2008); indeed they are two terms which often appear to be used to mean the same thing. One way to think of quality of life is as an umbrella term covering a variety of concepts such as: functioning, health status, perceptions, life conditions,

behaviour, happiness, lifestyle, etc. (Moons 2004). Sen defines well-being as being a narrower concept than quality of life and also introduces a third term, 'agency goals'. Well-being is used in the Capability Approach to refer to the individual's own physical and mental state (whether they are well nourished, happy, warm, etc). The individual's agency goals are not restricted to the advancement of the person's own well-being. Agency goals can be altruistic, so that the individual may value outcomes which benefit family, friends or society in general. The degree of freedom a person has to achieve their agency goals provides the most complete picture of their quality of life, but well-being will often be adequate for policy purposes (Nussbaum and Sen 1993).

The 1946 definition of health used by the World Health Organisation is so broad in its scope that with this definition it might be difficult to identify aspects of quality of life unrelated to health. Therefore, if it were health as "a state of complete physical, mental and social well-being" that was being measured comprehensively and incorporated into QALYs then it could reasonably be claimed that health is an adequate proxy for quality of life; this is not, however, the reality.

1.5 Using the Capability Approach to Assess Quality of Life in Patients with Chronic Pain

Each of the questions listed above which relate to how the Capability Approach should be operationalised has been addressed as part of this project. Through qualitative work involving patients with chronic pain information was gathered on how chronic pain impacts on quality of life. Framework analysis was used to identify nine higher-order capabilities from the qualitative data, along with a number of other capabilities which are related to and grouped under them. The list of higher-order capabilities was presented to and reviewed with participants involved in the original qualitative work and the final list was then developed into a questionnaire. The nine higher-order capabilities were used as the section headings for the questionnaire and within each of the nine sections were a number of related

questions. The respondent's answer to each set of questions indicates their level of achievement in terms of the higher-order capability. There were 42 questions in total. The questionnaire was initially piloted with twelve patients who were recruited through the Pain Management Clinic at the James Paget NHS University Hospital in Norfolk. This initial development of the questionnaire forms the first phase of the study and is reported in chapter eight.

The second phase of the study is outlined in chapter nine and involved administering the amended version of the questionnaire with a new group of pain patients, who were once again recruited through the Pain Management Clinic. This time the pilot involved 50 patients completing the questionnaire at three-monthly intervals over a total period of one year. Patients also completed EQ-5D and the EuroQol VAS and some simple comparisons are made between the results in chapter 10.

Finally, a multi-attribute value (MAV) method was chosen in order to elicit values for the nine higher-order capabilities and the levels within the dimensions/questions relating to each. The valuation exercise was completed by members of the public. A multiplicative functional form was chosen for the MAV model to allow the estimation of a score for each possible capability state defined by the questionnaire. Details of the MAV model, its origins and use in this study and results are found in chapter 11.

1.6 Summary of Results and Contributions to the Literature

Six focus groups were held, each comprising a distinct segment of the population. From the qualitative work nine higher-order capabilities were identified:

1. To be able to have self-respect
2. To be able to enjoy social interaction
3. To be able to fulfil the role of parent/ grandparent
4. Being able to remain physically and mentally active
5. Being able to have a positive and individual identity

6. Being able to be independent/ have control
7. Being able to participate in a loving relationship
8. Being able to enjoy physical and mental well-being
9. Being able to take enjoyment from life

The capabilities listed above incorporate those aspects of life discussed elsewhere in the literature on chronic pain, while having been conducted with the intention of identifying important capabilities. The capabilities listed above relate specifically to patients with chronic pain and the instrument developed in this study was developed to be condition specific. A limitation, therefore, is that the instrument is unlikely to be suitable for use with other patient groups and that if used in an economic evaluation relating to chronic pain then the results could not be compared with the evaluation of interventions for other medical conditions. The instrument will, however, be broad enough to detect changes in the quality of life of patients with chronic pain brought about by health, social and housing services.

What is more, the capabilities identified here are similar to those identified by Grewal, Coast *et al.* (2006) as being important to older people and those capabilities (adapted from Nussbaum's original list) which are being used in work by Lorgelly (2008) and Anand (Anand, Hunter *et al.* 2005). The similarity between these three lists is sufficient to suggest that one generic list of capabilities *could* be suitable and relevant for use as the evaluative space for economic evaluations across a range of UK policy contexts. After all, it is 'higher-order' aspects of quality of life which have been identified and included. It would seem a reasonable assumption that capabilities such as being able to achieve self-respect and enjoy social interaction will be important goals for most individuals, regardless of their specific health or social circumstances.

The higher-order capabilities and the levels within the questions were valued using a multi-attribute value method first used and reported by Peacock and Richardson (2007); from this preference values are inferred for the numerous capability states defined by the questionnaire. Chapter seven outlines the

methods and approaches which were available and the reasons for choosing the MAV method. The MAV method was found to be simple and useful and is recommended for wider use. It appears that part-whole bias was observed in the original results, but in an extension to the study outlined in chapter 10 it was found that steps could be taken to significantly minimise the effect of this bias. The result is significant as part-whole bias is usually associated with contingent valuation studies, and discussion of it so far has tended to be restricted to this context. The results of this study not only suggest that part-whole bias is a potential problem for other preference elicitation techniques, but that when one has an awareness of this steps can be made to remove the bias.

The thesis contributes to the relatively sparse literature on the Capability Approach in health economics. The approach brings the potential to assess broader quality of life and hence more of the dimensions considered to be important to patients. Although the focus of this study has been chronic pain, there is overlap between three lists of capabilities developed so far, in differing contexts. This overlap suggests that it may be possible and appropriate to develop one generic list for a broad range of evaluations. In turn, this would allow the cost effectiveness of interventions to be compared, even when these interventions span different policy contexts (health care, social services, housing services and so on). Capability therefore presents a feasible challenge to both cost-utility analysis and cost-benefit analysis, and would offer a broader and more transferable evaluative space than currently used in economic evaluation within health.

CHAPTER 2: Chronic Pain

This chapter begins by introducing the problem of pain, and more specifically what is meant by chronic pain. Section 2.1 will then provide an overview of how chronic pain affects quality of life. Section 2.2 introduces a range of outcome measures typically used to assess patients with chronic pain and section 2.3 summarises.

Pain has been defined as “an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage” (IASP, cited in Coniam and Diamond 1994, p2). It is necessary to be able to experience pain; pain tells us if a part of our body is being damaged so that we can avoid the cause (Gamlin and Lovel 2002). Indeed, Gamlin and Lovel state that in cases where pain is not experienced, due to conditions such as diabetes, damage can become widespread and serious before action is taken. This is not to say, however, that all pain is useful and when pain becomes repetitive or persistent then it is harmful to both physical and emotional well-being (Ibid p2). Coniam and Diamond report that in a large proportion of cases in which people seek treatment for pain there is no actual or potential tissue damage, and in these cases the pain is categorized as ‘useless pain’ (Coniam and Diamond 1994).

A distinction can therefore be made between acute pain (of short duration) which acts as a warning that damage is being done to the body, and chronic pain, which lasts for over six months, is ongoing, and is due to non-life threatening causes (Harris, Morley et al. 2003; Dysvik, Lindstrom et al. 2004). The causes of chronic pain are numerous and diverse, but include: nerve damage, rheumatoid arthritis, osteoarthritis, migraine headaches, traumatic injury, and surgery.

It is claimed that chronic pain affects 7.8million people in the UK of all ages (CPPC). Chronic pain is the second most common reason given by claimants

of incapacity benefit and it is estimated that the cost of back pain alone to the UK economy was as high as £12billion in 1998 (CPPC).

2.1 The effects of chronic pain on quality of life

Pain is tiring... it makes you want to snap out at the world. It's demoralizing, it's depressing, it colours your whole existence. It affects me day and night; it restricts everything and governs my life.

An anonymous Participant in an interview conducted by Carson & Mitchell (1998 p1244)

The above quotation provides a short but emotive introduction to sufferers' experiences of living with chronic pain and emphasises the degree to which chronic pain impacts on a great many aspects of a sufferer's quality of life.

This section is *not* intended to be a systematic review; instead the principle objective of 'dipping in to' the qualitative literature is to provide the reader with a *brief* introduction to chronic pain. Chronic pain can be defined in the clinical sense (and has been in the introduction to this chapter), but this does not give an indication of the human impact of the condition. Two systematic reviews (Breen 2002; Richardson 2005) have been published and they are adequate for our purposes here. The author of the first of these reviews notes how the issues reported by pain patients are consistent over time and regardless of the academic discipline of those conducting and reporting the qualitative work. Other articles are referenced on an *ad hoc* basis if they give a more in depth discussion of specific aspects of life, but however specific issues are discussed the picture is one of a list of broad themes which are common across the literature and summarised in both existing systematic reviews.

Following a review of the nursing, psychology and neurophysiology literature, Breen (2002) reports that there is a consensus across disciplines and decades that living with chronic pain alters life patterns resulting in negative physical, psychological, and social effects. Effects included alterations in physical patterns of eating, resting and sleeping, resulting in the need for inactivity and frequent periods of rest (*Ibid.* p54). Alterations in mobility result

in the loss of ability to perform the activities of daily living. Disability results in the loss of work, leading to economic stress and, in turn, resulting in strained, dysfunctional relationships with significant others. Psychological effects reported in the literature included: depression, anger, anxiety, grief, hopelessness, and feelings of helplessness. Alterations in social patterns resulted in isolation and loneliness (Ibid. p54).

In a separate study, conducted in Canada by Dewar, White *et al.* (2003, p46), sufferers of chronic pain were asked the question: 'what problems do you experience when dealing with your chronic pain?'. Responses were recorded and groups were then asked to select the most important. The process resulted in six priorities among the groups. In order of importance, these were: (1) Medical and treatment issues; (2) problems with activities of daily living; (3) emotional distress; (4) social issues; (5) sleep disturbance; (6) finance.

In reviewing the literature on chronic illness, Richardson (2005) notes that the importance of being believed is also often emphasised by study participants. Richardson reports how the invisible, subjective and contested nature of chronic pain leads to sufferers experiencing barriers to legitimisation of their condition.

From the two systematic reviews, a number of themes stand out. These themes are:

- The pain itself (intensity, frequency and location)
- medical treatment
- legitimisation and identity
- independence
- activities of daily living
- employment
- sleep disturbance
- relationships
- social issues and loneliness
- psychological and emotional distress

Respondents in the study by Dewar, White *et al.* (2003) reported experiencing difficulty finding accessible, effective and acceptable health-care, and several commented that they had experienced unacceptable attitudes and behaviours from their physicians. This included not being believed and receiving insensitive treatment and comments. Respondents also felt that they lacked knowledge about managing their condition and about making treatment choices.

In an article for *The Lancet*, Loeser and Melzack (1999) note that all types of chronic pain lead people to seek health care, but that they are often not treated effectively (p1609). Hadjimichael *et al.* (2007) have noted that within a group of patients with multiple sclerosis, respondents with severe pain made greater use of the healthcare system than those with moderate pain, but were less likely to be satisfied with their doctor's efforts to manage their pain.

It is reported to be difficult for chronic pain sufferers to express their pain and many go to extreme lengths in attempting to hide their pain and maintain a positive identity. Richardson (2005) found that, although participants in her study were invited to talk about their pain, they still felt a need to defend themselves against potential accusations of talking too much about pain. Participants attempted also to stress that they had a high degree of self-control and to maintain a positive identity (Ibid. p43). It has also been found that young people (aged up to 19 years) choose to hide their pain from their family and friends so as not to lose the affection of these people or lose participation in group activities (Carter, Lambrenos *et al.* 2002).

Werner, Isaksen *et al.* (2004) explore issues of self and shame in illness accounts from women with chronic pain. Werner, Isaksen *et al.* provide a summary of how they interpret participant's accounts of their own positive strength and their negative feelings towards talk of illness by others with similar pain:

We read the women's descriptions ... As a plot, their stories attempt to cope with psychological and alternative explanations of the causes

of their pain. As a performance, their stories attempt to cope with the scepticism and distrust they report having been met with. Finally, as arguments, their stories attempt to convince us about the credibility of their pain as real and somatic rather than imagined or psychological. (Werner *et al.* 2004 p1035)

Difficulties with activities of daily living reported by respondents in the study by Dewar, White *et al.* were: finding physical comfort; dressing and undressing; lifting; driving; doing housework and other recreational activities. Respondents felt that their lack of mobility increased their dependence on others. Of those who reported pain in a study by Thomas, Peat, *et al.* (2004), 58.7% reported that pain interfered with everyday life. Dysvik *et al.* (2004) found that 'role physical' was the most affected domain on the SF-36 when used with a sample of chronic pain patients; a domain which refers to limitations in daily activities. In a study conducted by Kerr, Fairbrother *et al.* (2004) respondents also reported the most profound effect upon the 'role physical', 'physical function' and 'social function' domains of SF-36.

Employment rates among the study sample of Kerr, Fairbrother *et al.* were much lower than the normal rate in Australia and in most cases unemployment was reported to be as a result of pain. Respondents aged 50 years and older, in a study conducted by Carson and Mitchell, described a "self-enforced withdrawal from usual activities and daily routines" (Carson & Mitchell 1998, p1245). One participant stated that: "The pain has totally ended my career because it was impossible for me to be at a job for 8 hours straight" (Ibid.). Employment was also found to be a significant issue for patients with diabetic painful neuropathy; approximately one-third of patients participating in a study by Tolle *et al.* (2006) reported some level of disruption in their employment status due to diabetic painful neuropathy.

Sayar, Arikan *et al.* (2002) have specifically focussed on sleep and chronic pain. According to Sayar *et al.*, it is estimated that between 50% and 88% of patients with chronic pain have significant sleep complaints.

Snelling (1994) has studied the effect of chronic pain on the family unit and states that chronic pain causes role tension, marital conflict, reduced sexual activity, and feelings of anger, anxiety, resentment and despondency in other family members. Chronic pain can affect the balance of household tasks within the family, as well as impinging on the parents' shared responsibilities related to child rearing. When parental contact with children fluctuates through a range of emotions, the relationship is negatively affected (Ibid.). In an interview conducted by Raheim and Haland (2006) one woman with fibromyalgia told how she was afraid that her partner found her lazy and how she felt guilty for not being able to provide more practical support for her family. In one study (Snelling 1994) the husband of a chronic pain sufferer stated that: "she feels more like a relative whom I'm fond of and like to look after, rather than a wife".

Kwan *et al.* (2005) found that only 3% of a group of 151 patients with chronic pain reported that their sexual function was largely unaffected by their pain.

Older people living with persistent pain and receiving care in their own homes or in sheltered accommodation described in a study by Blomqvist and Edberg (2002) how their pain was placing some form of strain on their relationship with carers and family, whether this was due to the effort of hiding one's true feelings, the fear of being perceived as a burden or 'bother', or the feeling of being treated unfairly or being misunderstood.

With respect to social issues, Dewar, White *et al.* report that respondents experienced difficulty interacting with friends, family and society. One participant commented that being unable to work made him more socially isolated and bored. Many commented that they had experienced discrimination and a lack of support (Ibid. p48). Snelling (1994) suggests that social contact with friends and relatives may be greatly reduced due to embarrassment on the part of the chronic pain sufferer, or simply as a consequence of self-preoccupation. The lack of predictability and ability to plan ahead also affects the sufferer's social life. One respondent from the study by Carson and Mitchell (1998) described vividly how: "sometimes I'm

getting ready to go out somewhere if my back isn't too bad, then all of a sudden it starts again and it's just like a cloud coming down. Where it was so bright now it's cloudy and it takes your energy away and so spoils everything" (P1244).

In a study conducted by Paulson *et al.* (2002) men with fibromyalgia reported becoming easily irritated, due to their continuous pain. Being off work the men experienced feelings of emptiness and shyness; they emphasised a contrast to before their pain when they felt that people had seen them as cheerful and kind, offering to help others if needed. They described how, during difficult periods, they became angry easily and this anger was directed towards their family. The men had given up hoping for a life without pain and were hoping instead simply that the pain would not get worse.

Several emotions were listed by respondents in the study by Dewar, White *et al.*, including: anger; depression; anxiety; irritability; upset feelings; limited patience (Ibid. p48). Mood swings and negative feelings contributed to disturbances in relationships and created high levels of stress. Participants expressed frustration at not knowing what had caused their pain. Kothe (2007) studied a group of patients with rheumatoid arthritis, some of whom also suffered from low back pain and found that those patients with arthritis and low back pain had significantly higher depression scores compared to patients with arthritis but no low back pain. As noted earlier, Breen (2002) lists the psychological effects of chronic pain described in the literature as including: depression, anger, anxiety, grief, hopelessness, and feelings of helplessness.

To summarise, it can be seen that chronic pain has a broad and significant impact on quality of life. Sufferers are restricted in their ability to be independent, to work and to go about daily activities. There is a negative impact on roles and relationships, on sleep and emotions, and in terms of the general enjoyment of life. Sufferers struggle to maintain a positive identity, to legitimise their condition and in seeking effective treatment. Practically, they

are a patient group who relies heavily on assistance from friends and family, social services, and health professionals.

2.2 Common Outcome Measures for Chronic Pain:

Table 2.1 introduces four disease-specific instruments designed specifically for use with pain patients. The scope of these instruments will be considered, in terms of the type of dimensions they include, and relative to a broad definition of health given by the World Health Organisation.

Table 2.1: Common Outcome Measures Used In Chronic Pain

Measure:	Brief Pain Inventory (BPI)	Author:	Charles S. Cleeland
Description:	<ul style="list-style-type: none"> ▪ Developed to provide information on the intensity of pain; degree to which pain interferes with function; pain relief; patient's perception of the cause of pain. ▪ Initially assesses pain intensity at its worst, best and at the time of the evaluation using a numerical rating scale (0-10). ▪ Using numerical 0-10 scales, the BPI then asks for ratings of the degree to which pain interferes with mood, walking, and other physical activity, work, social activity, relations with others, and sleep. ▪ The mean of these scores can be used as a pain interference score. ▪ Further assesses percentage of relief from current medications or treatments, and duration of relief 		
Comments:	<ul style="list-style-type: none"> ▪ Full 20-item BPI takes about 10-15 minutes to complete. 		
Measure:	Short-Form McGill Pain Questionnaire	Author:	Roland Melzack
Description:	<ul style="list-style-type: none"> ▪ The questionnaire was designed to provide quantitative measures of clinical pain ▪ Respondent selects a column (e.g. by ticking) according to the degree to which they feel a particular type of pain, so for example they would mark according to the degree to which they feel "Throbbing", "Shooting", or "stabbing" Pain ▪ For each type of pain the respondent can indicate that the degree of pain is "none", "Mild", "Moderate", or "Severe" ▪ Also a present pain intensity scale (visual analogue scale). 		
Comments:	<ul style="list-style-type: none"> ▪ The original questionnaire is difficult for many people to complete ▪ Original questionnaire takes 10-15 minutes to administer, short form takes 2-5 minutes to administer. 		

Measure:	Oswestry Disability Index (ODI)	Author:	Fairbank, J
Description:	<ul style="list-style-type: none"> ▪ Specific to back pain ▪ Indicates the extent to which a person's functional level is restricted by pain. ▪ Consists of 10 items: Pain intensity; Personal care; Lifting; Walking; Sitting; Standing; Sleeping; Sex life; Social life; Travelling. ▪ There are six levels for each, specifically detailing the degree of restriction 		
Comments:	<ul style="list-style-type: none"> ▪ Self-administered ▪ Takes 5 minutes to complete 		

Measure:	Chronic Pain Grade (CPG)	Author:	M von Korff J Ormel FJ Keefe SF Dworkin
Description:	<ul style="list-style-type: none"> ▪ The questionnaire is proposed as an interview-administered, multi-dimensional measure of chronic pain severity. ▪ The questionnaire has also been administered in a postal survey in the UK. ▪ Contains seven questions. ▪ Pain severity at present, intensity, pain interference, and ability to take part in activities of daily living over a six month period are all assessed using a scale of 0-10. The number of days on which the respondent has been kept from their usual activities is also recorded. ▪ There is a scoring system in which points are assigned to questions. ▪ Classifies respondents into one of five hierarchical pain grades: pain free (grade 1), low disability and low intensity (grade 2), low disability, high intensity (grade 3), high disability, moderately limiting (grade 4) and high disability, severely limiting (grade 5) 		
Comments:	<ul style="list-style-type: none"> ▪ From research conducted in the UK the CPG was found to have good internal consistency and reliability. Convergent validity and construct validity were also confirmed. 		

Authors as listed on <http://www.proqolid.org/>, with the exception of the Chronic Pain Grade, see Von Korff, Ormel, *et al.* (1992).

The outcome measures listed in Figure 2.1 are narrow in their focus and they are not intended for use as measures of health-related quality of life. In keeping with the World Health Organisation's 1946 definition of health as total social, psychological and physical well-being, Bowling (1997) reports that measures of health-related quality of life are typically expected to include such dimensions as physiology, function, social activity, cognition, emotion, sleep and rest, energy and vitality, health perception, and general life

satisfaction. It is appropriate to think of health-related quality of life as a broad consideration of social, psychological and physical well-being, viewed from the patient's perspective. It should be realised, however, that health-related quality of life is a concept seldom defined. Furthermore, it should not be expected that every measure of health-related quality of life will cover each and every one of the nine dimensions listed above, or indeed place equal emphasis on each.

The McGill Pain Questionnaire focuses entirely on the severity and nature of pain. The Oswestry Disability Index, Brief Pain Inventory and Chronic Pain Grade, although still disease-specific measures for pain, all include a slightly wider range of dimensions. As can be seen from Table 2.1, the ODI has ten dimensions, mainly concerning limitations to physical functioning (walking, sitting, standing, personal care, and lifting) but there are also dimensions relating to social life, sex life and travelling. Among others, the BPI contains dimensions relating to general activity, normal work, relations with other people and enjoyment of life. The CPG contains dimensions relating to interference with daily activities (work, school or housework) and recreational, social and family activities. Given how broadly health is defined by the WHO, and the equally broad definition of health-related quality of life offered by Bowling (1997), it is not at all clear whether the ODI, BPI and CPG can or should be classed as measures of health-related quality of life. What is certain is that as disease specific measures these three instruments would be considered unsuitable under NICE guidelines for use in a cost-utility analysis. Guidelines state that health states should be described using a "standardised and validated *generic* instrument" (NICE 2008, p21. Emphasis added).

Figure 2.1: Questions from the Nottingham Health Profile

<p>Physical Abilities</p> <ul style="list-style-type: none"> ▪ I can walk about only indoors ▪ I find it hard to bend ▪ I'm unable to walk at all ▪ I have trouble getting up or down stairs or steps ▪ I find it hard to reach for things ▪ I find it hard to get dressed by myself ▪ I find it hard to stand for long ▪ I need help to walk about outside
<p>Social Isolation</p> <ul style="list-style-type: none"> ▪ I feel lonely ▪ I find it hard to make contact with people ▪ I feel that there is nobody that I am close to ▪ I feel I am a burden to people ▪ I'm finding it hard to get along with people
<p>Sleep</p> <ul style="list-style-type: none"> ▪ I take pills to help me sleep ▪ I'm waking up in the early hours of the morning ▪ I lie awake for most of the night ▪ It takes me a long time to get to sleep ▪ I sleep badly at night
<p>Emotional Reaction</p> <ul style="list-style-type: none"> ▪ Things are getting me down ▪ I've forgotten what it's like to enjoy myself ▪ I'm feeling on edge ▪ The days seem to drag ▪ I lose my temper easily these days ▪ I feel as if I'm losing control ▪ Worry is keeping me awake at night ▪ I feel that life is not worth living ▪ I wake up feeling depressed
<p>Pain</p> <ul style="list-style-type: none"> ▪ I have pain at night ▪ I have unbearable pain ▪ I find it painful to change position ▪ I'm in pain when I walk ▪ I'm in pain when I'm Standing ▪ I'm in constant pain ▪ I'm in pain when going up or down stairs ▪ I'm in pain when I'm sitting
<p>Energy Level</p> <ul style="list-style-type: none"> ▪ I'm tired all the time ▪ Everything is an effort ▪ I soon run out of energy

(Source: <http://www.cebp.nl/media/m83.pdf>)

The Medical Outcomes Study Short-form general health survey-36 (SF-36) and the Nottingham Health Profile are two generic measures of health-related quality of life which have commonly been used with chronic pain patients.

Both are self-administered and both take approximately 5-10 minutes to complete (www.proqolid.org). The Nottingham Health Profile was developed in the UK and aims to provide a brief indication of a patient's perceived emotional, social and physical health problems (www.proqolid.org). The Nottingham Health Profile is not an index of disease, illness or disability but relates instead to how people feel when they are experiencing various states of ill health (Bowling 1997, P44). The questions (listed in Figure 2.1) are answered by the respondent with a simple 'yes' or 'no'.

The SF-36 includes eight domains and is the most widely used generic measure of health in clinical trials. The SF-6D is derived from the SF-36, specifically for use in economic evaluation and contains six dimensions: physical functioning; social functioning; role limitations; mental health; energy/vitality; pain. The SF-36 allows Quality Adjusted Life Years to be obtained for use in cost-utility analysis. Questions from the SF-6D can be found in Figure 2.2.

A number of issues have been encountered within this section which will be addressed more fully in the following chapter. Specifically, these issues are: (i) a lack of clarity regarding the concept of health-related quality of life; (ii) the apparently all encompassing definition of health set by the WHO; (iii) the type of outcomes (i.e. aspects of life) which are included in cost-utility analyses and the suitability of these.

Figure 2.2: SF-6D (Drummond, Sculpher et al. 2005)

Physical Functioning <ol style="list-style-type: none">1. Your health does not limit you in vigorous activities2. Your health limits you a little in vigorous activities3. Your health limits you a little in moderate activities4. Your health limits you a lot in moderate activities5. Your health limits you a little in bathing and dressing6. Your health limits you a lot in bathing and dressing
Role Limitations <ol style="list-style-type: none">1. you have no problems with your work or other regular daily activities as a result of your physical health or emotional problems2. You are limited in the kind of work or other activities as a result of your physical health3. You accomplish less than you would like as a result of emotional problems4. You are limited in the kind of work or other activities as a result of your physical health and accomplish less than you would like as a result of emotional problems
Social Functioning <ol style="list-style-type: none">1. Your health limits your social activities none of the time2. Your health limits your social activities a little of the time3. Your health limits your social activities some of the time4. Your health limits your social activities most of the time5. Your health limits your social activities all of the time
Pain <ol style="list-style-type: none">1. You have no pain2. You have pain but it does not interfere with your normal work3. You have pain that interferes with your normal work a little bit4. You have pain that interferes with your normal work quite a bit6. You have pain that interferes with your normal work extremely
Mental Health <ol style="list-style-type: none">1. You feel tense or downhearted or low none of the time2. You feel tense or downhearted or low a little bit of the time3. You feel tense or downhearted or low some of the time4. You feel tense or downhearted or low most of the time5. You feel tense or downhearted or low all of the time
Vitality <ol style="list-style-type: none">1. You have a lot of energy all of the time2. You have a lot of energy most of the time3. You have a lot of energy some of the time4. You have a lot of energy none of the time

The remainder of this chapter therefore judges outcome measures simply according to whether they appear to assess and incorporate those aspects of quality of life found in section 2.1 to be important to patients with chronic pain. This exercise is summarised in Table 2.2. Recall from the previous section that eleven themes were identified from the literature on chronic pain. Medical treatment was a theme which incorporated considerations such as the attitude of the physician and the patient's understanding of their condition. Medical treatment is excluded from Table 2.2 as it is not felt that an

intervention can reasonably impact on these factors and because effective treatment should be expected to cause improvement in the other aspects of life, meaning that to include both in an outcome measure risks double counting benefits. Furthermore, access to medical treatment will depend on the system of healthcare provision in place. Some dimensions listed in Table 2.2 may only partly or tenuously be included in the instruments assessed; how fully they are included is largely a subjective judgement.

The Brief Pain Inventory appears to incorporate more of what is important to patients with chronic pain than other instruments, as it has been specifically developed for use with pain patients. The BPI is not an instrument that would typically be used to obtain QALYs however. From the generic health-related quality of life measures, the Nottingham Health Profile appears to include more important aspects of life than the EQ-5D or SF-6D. The SF-6D is the worst instrument in terms of incorporating what is important to patients with chronic pain. Just four of the 10 aspects of life identified from the literature are fully included; in terms of independence, the scope of SF-6D is restricted to being able to bath and dress oneself.

Table 2.2 Inclusion of Dimensions Important to Patients with Chronic Pain

Dimension/ aspect of quality of life	Outcome Measure						
	EQ-5D	CPG	SF-6D	Nottingham Health Profile	ODI	BPI	Short-form McGill
Pain intensity and/or frequency	✓	✓	×	✓	✓	✓	✓
Legitimation and identity	×	×	×	×	×	×	×
Activities of daily living	✓	✓	✓	only physical functioning	✓	✓	×
Independence	Self-care	×	Bathing & dressing	Getting dressed	Personal care	×	×
Employment	✓ If included under usual activities	×	✓ Interference with normal work	×	×	✓	×
Sleep disturbance	×	×	×	✓ Sleep & energy levels	✓	✓	×
Relationships (and family)	×	✓ Interference with family activities	×	✓ Partly included in social isolation	Limited to sex life	✓	×
Social issues	✓ If included under usual activities	✓	✓	×	✓	✓ If included in general activity	×
Loneliness	×	×	×	✓ Social isolation	×	×	×
Psychological and emotional distress	✓	×	✓	✓	×	✓ Mood	×

2.3 Summary

The studies discussed in this section give an overview of the problem of chronic pain, demonstrating how pain affects all aspects of an individual's life. Richardson summarised that participants with widespread pain reported in interviews that: "the chronicity of their pain had transformed their world into one in which pain had become a normal part of everyday life and previous 'normal' life could no longer be brought to mind" (Richardson 2005, p35).

Health-related quality of life is a term which is rarely defined. Health, as defined by the World Health Organisation, and health-related quality of life, as interpreted by Bowling from the WHO definition, are actually very broad concepts. Bowling suggests that dimensions of health-related quality of life can include: physiology, function, social activity, cognition, emotion, sleep and rest, energy and vitality, health perception, and general life satisfaction. None of the seven instruments considered in this chapter contain all of the dimensions suggested by Bowling, although the Nottingham Health Profile appears to be the most comprehensive measure of health-related quality of life. There appears to be a slight mismatch between how health-related quality of life has been interpreted conceptually (an interpretation which is actually very broad, encompassing most perceivable aspects of quality of life), and the scope of health-related quality of life measures (which tend to contain a number of dimensions on health and physical functioning, alongside one or two rather general questions relating to social activity/ participation).

It seems clear, however, that the impact of chronic pain extends beyond the more limited interpretation of health-related quality of life adopted by instruments such as the SF-6D and EQ-5D, which are used to generate QALYs for use in economic evaluation. Wider areas of an individual's quality of life that may be affected by chronic pain include: the struggle by sufferers to find legitimisation of their condition and to maintain a positive identity; sleep disturbance; social isolation and loneliness; relationships and family; general enjoyment of life. Most instruments include quite broad dimensions relating to activities of daily living (which can include employment) and mood/emotional distress, but it is unclear to what extent these 'catch-all' phrases encompass those 'missing' or 'extra' dimensions listed above. Indeed there

may be variation in how (broadly) patients interpret the dimension 'daily activities' and what they consider they should be able to do in terms of daily activities may be influenced by adaptation.

The following chapter will explore how changes in health-related quality of life, as assessed by instruments such as those introduced in this chapter, are incorporated within economic evaluations and will challenge the appropriateness of this.

CHAPTER 3: Measuring Outcomes for Economic Evaluation

3.1 Introduction:

It is widely accepted that health care should be classed as an economic good, that is to say that the resources used to produce health care are finite and devoting more resources to the treatment of a particular patient group results in resources being diverted away from some other use. Other uses for the same resources may include the treatment of other patient groups, or alternatively it may be the production and consumption of goods in entirely different areas of the public or private sectors. No health care system, anywhere in the world, has achieved levels of spending sufficient to meet all its clients' wants for health care (Morris, Devlin et al. 2007, p3).

The scarcity of the resources required for health care requires economic agents to make choices. In the UK, given the existence of the National Health Service, it is most typical to think of citizens as making decisions over whether to consume health services which are free at the point of consumption. This is not to say that there will be no opportunity cost associated with the consumption, or that patients behave exactly as they would when consuming private goods. For example, it tends to be the medical profession who advise the consumer which treatments are necessary and available (the agency relationship). The availability of different treatments from the National Health Service is determined by policy makers who are likely to consider both efficiency and equity as well as a variety of other concerns (e.g. affordability, lobbyists, political expediency). The framework used by economists to inform policy makers in this respect is economic evaluation, an umbrella term for a range of specific approaches.

Efficiency can broadly be defined as "the allocation of scarce resources that maximises the achievement of aims", i.e. benefits (Knapp (1984), cited in Morris, Devlin et al. 2007). So, if only benefits are being considered, then the objective of public policy should be defined in terms of the maximisation of these benefits (Dolan and Robinson 2001, p1698). However, policy makers as well as the public are also likely to be concerned with how benefits are distributed (Ibid.). Equity is largely synonymous with the fairness of a particular distribution and doesn't necessarily

imply that the distribution is equal. It is common to talk of an 'equity-efficiency trade-off' (Reinhardt 1998) and a social-welfare function seeks to show the trade-off between the maximisation of benefits and an equitable distribution of those benefits (Dolan and Tsuchiya 2009). The nature (shape) of the social-welfare function falls outside the scope of this thesis, but the distinction between efficiency and equity is important and will be discussed in more detail section 3.3.2.

Economic evaluation can be defined as: "the comparative analysis of alternative courses of action in terms of both their costs and consequences" (Drummond, Sculpher et al. 2005, p9). Therefore, the basic tasks of any economic evaluation are to identify, measure, value and compare the costs and consequences (often referred to as outcomes or benefits) of the alternatives being considered.

The identification of various types of costs and their subsequent measurement in monetary units is similar across most forms of economic evaluation. The principal focus of this thesis is the identification, measurement and valuation of the consequences stemming from the alternative interventions being considered. The nature of the inclusion of consequences differs considerably between three common types of economic evaluation (Drummond, Sculpher et al. 2005).

Cost-effectiveness analysis (CEA) is used when the consequences of different interventions may vary but can be measured in identical natural units. Competing interventions are compared in terms of cost per unit of consequence. Examples of units of outcome that could be used in CEA are blood pressure, survival rates or levels of cholesterol.

The second option is Cost-Utility Analysis (CUA), in which consequences of interventions are compared in terms of utility. CUA is a broader form of analysis than CEA, but is a variant of that general approach. In the case of CUA, changes in quality of life and changes in the quantity of life are commonly combined in order to present the outcome in terms of quality-adjusted life years (QALYs). Competing interventions are therefore compared in terms of cost per QALY.

Cost-benefit analysis (CBA) is a third option, in which both costs and consequences of different interventions are expressed in monetary units.

Cost-minimisation analysis (CMA) can be used when it can be demonstrated that the consequences of the interventions being considered are the same. In this case only inputs (costs) need to be taken into consideration. The aim here is simply to find the cheapest way of achieving the same outcome.

The different types of economic evaluation are summarised in Table 3.1.

Table 3.1: Measurement of consequences in economic evaluation

Type of study	Identification of consequences	Measurement/ valuation of consequences
Cost-minimisation analysis	Evidence of identical consequences	None
Cost-effectiveness analysis	Single effect of common interest, common to both alternatives, but achieved to different degrees	Natural units
Cost-utility analysis	Single or multiple effects, not necessarily common to both alternatives	Quality-adjusted life years
Cost-benefit analysis	Single or multiple effects, not necessarily common to both alternatives	Monetary units

Adapted from: (Drummond, Sculpher et al. 2005, P2)

Leaving aside cost-minimisation, the remaining three approaches to economic evaluation can broadly be categorized under one of two headings: Welfarist and Non-Welfarist. Cost-benefit analysis will be considered in Section 3.2 as a welfarist approach to economic evaluation. Cost-effectiveness and typically cost-utility analysis are non-welfarist approaches and will be considered in section 3.3. The discussion surrounding each approach will include the issue of how equity is dealt with.

The two main differences between what is welfarist and what is non-welfarist is the evaluative space in which states of the world are assessed and the perspective from which states of the world are valued. The evaluative space will therefore be covered in more detail in Section 3.4 and the perspective for valuation will be discussed more

fully in Section 3.5. Section 3.4 will also introduce a further approach, experienced utility, which has only recently attracted attention in health economics but which is firmly rooted in the utilitarian tradition.

Section 3.6 concludes by addressing the question which, indeed if any, of these methods would appear to be the most appropriate for use in an economic evaluation of different interventions to relieve chronic pain?

3.2 Welfarist Approach

Although it is possible for cost-benefit analysis (CBA) to be non-welfarist (Brazier, Ratcliffe et al. 2007), the form of cost-benefit analysis outlined here will conform with welfarist tenets. The four main tenets of conventional welfarist practice (Brouwer, Culyer et al. 2008) are:

- I. Individuals rationally maximise their welfare (expressed as utility)
- II. Individuals are themselves the best judges of what contributes most to their utility and how much that contribution is
- III. Utility is only derived from the outcomes of behaviour and processes
- IV. The goodness of any situation should be based solely on the basis of utility.

The first tenet implies, and the fourth states explicitly, that it is utility (generally thought of as happiness and satisfaction) which is the appropriate metric for assessing the goodness of different states of the world, and the second states that it is individuals (consumers or households) who are the best judges of which state they would rather be in. Welfare economics is based entirely on individualism: the only relevant information in making social choices is the views of the individuals affected by those choices. The third tenet is significant because it states that utility is only generated from the consumption of a good or the outcome of a process (such as health treatment) and not from other aspects, such as obtaining the good or from the process which yields the outcome. Welfarism can therefore be said to be consequentialist.

Welfarists suggest that individuals should use utility as the metric with which to value the goodness of different states of the world and that what they are valuing is any

outcomes or consumption associated with a particular state. Each state can be ranked according to the amount of utility gained in that state and the individual decision maker will choose the option leading to the state in which the level of utility is greatest. Problems arise when we seek to aggregate individual utilities and the policy-maker is then confronted with situations where the relative desirability of states depends upon what sorts of trade-offs between individual utilities are deemed acceptable (Morris, Devlin et al. 2007, p211).

In welfare economics, this problem is partly overcome by relying on Pareto's Law: we can say that a new state of the world, or outcome, (B) is better than another (A), if some people are better off in state – or with outcome – B, while at the same time no person is worse off than they were in A. An outcome is Pareto Optimal if an increase in any one person's utility can only be achieved by reducing the utility of at least one other person (Morris, Devlin et al. 2007, pp211-220).

In turn, at least two criticisms can be made of Pareto optimality. First, it provides us with little guidance about distribution, starting with the status quo, and ignoring inequalities. Second, numerous states or outcomes can be Pareto improvements, and we cannot rank these. On the other hand, there may only be cases where some gain and some lose and equally we cannot rank these either (Morris, Devlin et al. 2007). In response to these weaknesses came the Kaldor-Hicks criterion, which factors in the possibility of compensating losers.

Given the fact that we have no dependable and cardinal measure of utility – there are no numbers on a measuring rod representing the same level of utility for all individuals – economists have instead tended to use money as a proxy for utility. Hence, cost-benefit analysis involves measuring both the inputs and consequences of different interventions in monetary units (Ibid.).

The most common approach to the monetary valuation of health outcomes is willingness-to-pay (WTP), through the technique of contingent valuation. In recent years there has been a rapid growth in the number of contingent valuation studies published in the health care literature (Sach, Smith et al. 2007), With the review by Sach *et al.* identifying 210 studies that were undertaken in high income OECD

member countries. Although the search covered the period from 1985, many of the articles identified were published between 2000 and 2005. There are also many studies which have been experimental in nature, attempting to explore measurement feasibility issues (Drummond, Sculpher et al. 2005).

There have been three primary arguments advanced for the superiority of WTP over the cost-utility approach (Olsen and Smith 2001; Sach, Smith et al. 2007). First, it is suggested that WTP imposes no restriction on which attributes of a programme people are allowed to express a value for. Second, WTP is deemed to be the theoretically correct approach, because of its foundation in welfare economics. Third, WTP values benefits in the same unit as costs.

The first of the three arguments for the superiority of cost-benefit analysis was that CBA has the potential for a broader scope in its valuation of benefits. The argument here is that cost-utility analyses assess preferences for health state improvement only, which represents a focus on a single influence on quality of life. Both illness and health interventions may have some impact on elements of quality of life other than just health. The health and non-health effects can be valued using WTP and hence the full impact is estimated. The inclusion of both health and non-health impacts and the use of money as a common metric also mean that health interventions can be evaluated alongside interventions from outside of the health sector, whether this be education initiatives, road safety or environmental policy. This is the key strength of WTP, although – as will be seen in the following paragraphs - there are theoretical and practical issues associated with the approach.

The theoretical basis for the use of WTP, as already stated, is the Kaldor-Hicks (or potential Pareto improvement) criterion. One implication stemming from the Kaldor-Hicks criterion is that it can lead to decisions which favour the rich at the expense of the poor because the rich have greater ability to pay in support of any given strength of preference (Donaldson 1999). The result of using WTP when applied in a strictly welfarist form would then be to produce greater expenditures to protect wealthy groups than to protect the poor (Sunstein 2001), a result that would surely be controversial. Chronic pain patients would be a group which would almost certainly lose out from such a result, as they are more likely to experience regular and/or

prolonged periods of absence from employment, resulting in many cases in some degree of financial hardship. A practical solution to this particular issue is to elicit WTP from members of the general public³ and to use the average of the WTP across the sample group to inform policy.

One method of significantly limiting the impact of budget constraints, and a more realistic way of presenting WTP questions is to elicit respondents' WTP to avoid some probability of becoming ill in the future, or alternatively WTP for insurance to guarantee them treatment should they become ill. The problem, however, then becomes one of making the task more cognitively demanding for respondents by including probabilities. Research by Hammitt and Graham (1999) demonstrates that WTP for risk reduction is inadequately sensitive to the differences in the probability. The authors suggest that in many cases poor study design may largely be to blame for the insensitivity of stated valuations to changes in risk magnitude.

A further problem relating to equity is that the Kaldor-Hicks criterion only requires that there is the *potential* for losers to be compensated, and this is unlikely to happen in reality. If compensation were to happen then the facilitation of this compensation will add in an additional cost.

Ability to pay creates additional problems in the respect that the WTP reported by the poor for a particular health intervention may only crudely be related to their valuation of the outcome of that intervention (Sunstein 2001, p254). Poor people are able to pay less than wealthy people simply by virtue of being poor, even though it is possible that they are willing to pay a higher percentage of their wealth. This is not, therefore, an accurate reflection of the strength of desire or of expected happiness. We can therefore question the accuracy and usefulness of the reported monetary valuation as a proxy for expected utility. Empirical evidence showing insensitivity of WTP values to changes in the scale of the outcome appears to support this hypothesis. Findings by Smith (2005) suggest that insensitivity may be significantly determined by the individual's budget constraint so that the higher the proportion of

³ In reality samples of 'the general public' may not be truly representative as in addition to self-selection (where certain people are more willing to volunteer to take part in studies), certain groups such as prisoners and the homeless are frequently excluded due to practical difficulties associated with recruitment and participation.

income the expressed WTP represents the greater the insensitivity of that WTP to changes in the scale of the good, irrespective of changes in underlying marginal utility.

There have certainly been cases where a substantial discrepancy has been found between the ranking of programmes implied by respondents' WTP values and their explicit stated ranking of the programmes (Shackley and Donaldson 2002). What this observed preference reversal means is that WTP values do not always reflect the stated preference of respondents and this in turn casts doubt on the use of WTP to aid priority setting (Ibid. p973). The reason for this preference reversal is not fully understood and it may be a reflection of the fact that either explicit ranking involves no sacrifice, but WTP involves having to give up some amount of money, or that respondents adopt more of a societal view when doing the explicit ranking.

The majority of health services are private goods (they can be individually bought and sold). It can therefore be deemed appropriate that WTP questions be framed in such a way that payment would be made by the individual and that their valuation reflect the benefit they receive from the good or service they are paying for. However, it has been noted that under a national health scheme, the payment (usually through taxation) is largely unrelated to the benefits that the payee will obtain (Smith and Richardson 2005a, p79). There may be differences between the choices people make as consumers and the choices that they make as citizens, and it is not clear that the former should be preferred. The context of citizenship may evoke altruistic values that are not reflected in private choices and individuals may care about equity when it comes to priority setting at a societal level.

It can be questioned whether it is appropriate to think of 'social' WTP as simply the sum of 'individual' WTP. The relevant WTP question might be one of how much each individual is prepared to contribute to a central budget from which other people's treatment is paid for, but which would also cover the cost of treatment for the individual should they need it (Smith and Richardson 2005a). The danger is then that if respondents believe that a public policy-maker is using WTP values as evidence on which to base decisions as to whether or not to provide a new 'free' service, they might strategically over-state valuations to increase the likelihood of the

service being provided (Whynes, Frew et al. 2003). Conversely, Sunstein (2001) also suggests that people care about their relative economic position, not simply their absolute economic position, and hence they would be willing to pay more if they could be assured that others would be paying as well; when everyone is paying, people can maintain their relative economic position while also receiving the benefit.

Discussion so far has focussed on theoretical issues; there are also a number of evident and perceived practical difficulties with the WTP approach which will now be outlined. They may not be fundamental weaknesses, but at best they are potential pitfalls which make it essential to take great care with study design. The first is the perception that attaching monetary values to non-marketed goods and services will be intrinsically alien to respondents (Whynes, Frew et al. 2003). A similar perception is that respondents may feel deeply uncomfortable about attaching monetary values to health and will thus record 'protest' valuations of zero (Tsuchiya and Williams 2001; Coast 2004).

For example, in 1998 a questionnaire was used in the South Eastern region of France as part of the 'EuroWill' project which asked respondents to value three projects: heart operations, breast cancer treatment, and a helicopter ambulance service (Protiere, Donaldson et al. 2004). The number of zero responses interpreted as protest votes were 32%, 30% and 40% for each of the three projects respectively. As part of the same project, a questionnaire was used with respondents in the Republic of Ireland during 1999. Here the three programmes valued were: cancer treatments, heart operations and community care services for the elderly (Shackley and Donaldson 2002). In this study the percentage of zero protest votes were 10%, 12% and 15% for the three projects respectively.

Part-whole bias is a further and commonly observed practical difficulty with cost-benefit analysis. Part-whole bias is generally thought to occur when "respondents are unable to differentiate between benefit subcomponents or between the subcomponents and the value for all types of benefits" (Mitchell and Carson 1989, p251). In short, the respondent confuses some broader or wider reaching entity with the entity that the researcher wants to value. What is seen, therefore, is that respondents state much the same willingness to pay for different scales of

projects/benefit; this phenomenon has also been termed 'embedding' (Svedsater 2000). For example, Currie *et al.* (cited in Baker, Robinson *et al.* 2008) found an insensitivity in WTP responses to bring about two expansions in a cancer programme that would treat either 300 or 450 additional patients.

Suggestions can be found in the literature as to how to reduce the extent of part-whole bias. Among general strategies are: inclusion of the larger entity in the scenario, with warning not to confuse the larger entity with that being valued; making the description of the entity more salient, for example through the use of visual aids; have them first value the total resource, even if this is not the subject of the study, and then have them allocate their total WTP amounts for the individual components (Mitchell and Carson 1989).

Objections to the use of traditional welfare economics in the area of health have also been raised on the grounds of the specific characteristics of health care as a commodity. Culyer (1991) states that: "implicit in [the welfarist] approach is the presumption that goods and services are in principle best allocated by market mechanisms and that departures from that mechanism require special warrants" (p69). He then goes on to list several reasons why we may consider health care to be different from other goods and services and suggests that because of these differences we may wish to allocate health care differently and evaluate the efficiency of different allocations differently. One difference which Culyer highlights is the competence of the consumer. The argument here is that welfare economics assumes a degree of knowledge on the part of consumers which in the case of health many consumers do not have. Instead it is common for the consumer to defer judgements to a medical doctor, who acts as an agent on his or her behalf (Culyer, 1991). This argument is linked to the practical issue already highlighted, in that most consumers are just not familiar with valuing or being asked to pay for health care beyond the fixed cost of a prescription and treatment at the dentist.

To summarise, the evaluative space for welfare economics is utility. With cost-benefit analysis, money is used as a proxy for utility, which cannot be reliably measured on a cardinal scale. Three advantages of cost-benefit analysis are suggested: the potentially broad evaluative space, its theoretical foundations, and

the fact that both inputs and consequences are measured in monetary units. The broad evaluative space and the use of money to measure outcomes mean that health interventions can potentially be evaluated alongside any other public or even private sector project and this is the key strength of CBA. It is the case however that the Kaldor-Hicks criterion (the theory underpinning CBA) has very controversial implications for equity, and money may in fact not be a good proxy for utility. Furthermore, the fact that respondents have no experience of valuing the non-marketed good that is health care in monetary terms and may well object to doing so is not the only practical pitfall to consider and overcome when designing a willingness to pay study. Problems with WTP such as those set out above led to the development of CEA as an alternative form of evaluation for health care.

It will be seen in Section 3.4.3 that the broad evaluative space, which is the major strength of CBA, has not, in reality, been exploited so far in the context of health.

3.3 Non-Welfarism

3.3.1 Cost-Effectiveness Analysis

A theoretical basis for cost-effectiveness analysis can be found in the theory of production efficiency (Morris, Devlin et al. 2007). CEA tries to identify cases where more benefit can be produced at the same cost or where a lower cost can be achieved but with equal benefit. When using CEA we should already have decided that the outcome is worth doing, we are merely trying to find the most efficient way of achieving that outcome. With CEA, the type of outcomes we are able to consider are quite restricted. As mentioned in the introduction to this chapter, consequences must be expressed in terms of a single dimension, such as life years gained or number of episode free days. Even where this is practical to consider consequences in terms of a single natural unit (such as years of life), it immediately and very severely restricts the evaluative space.

It would be possible, for example, to assess the cost-effectiveness of a programme in terms of the cost of reducing a patient's self-reported pain score (measured on a rating scale) by one unit. There are two problems, however, with this idea. First,

pain is highly subjective. Second, other aspects of life which are likely to be affected by the programme (and which are discussed in chapter two) are effectively ignored.

Should CEA be used, the decision-maker would also be unable to consider the opportunity cost of funding an intervention to relieve chronic pain as opposed to funding an intervention from another area, unrelated to chronic pain, from the same limited health care budget. This is because the primary measure of effectiveness is likely to differ across different types of intervention.

3.3.2 Cost-Utility Analysis

Commonly within health economics a non-welfarist approach is being adopted in which the evaluative space is essentially limited to health, or *health-related* quality of life (Birch and Donaldson 2003; Mooney and Russell 2003; Coast 2004; Coast 2009). This narrow focus on health and the assertion that what matters to decision makers and to society is the maximisation of health, has been labelled by some as the 'decision maker's' approach (Culyer 1991; Brouwer and Koopmanschap 2000; Coast 2004).

What CUA generates is a set of (average) weights, often referred to as utilities, one for each possible health state, which reflect the relative desirability of each of those health states (Drummond, Sculpher et al. 2005). The utility weights for each health state are treated as cardinal values which lie on a scale that is established by assigning a value of 1 to being healthy and 0 to being dead (Torrance 1986), minus values are also allowed.

Quality adjusted life years (QALYs) are then calculated by multiplying the utility weights reflecting the relative desirability of a health state by the amount of time spent in that health state, summing over all time periods and standardising to a year (Morris, Devlin et al. 2007).

Various techniques such as the rating scale, standard gamble and time trade-off are used to elicit utility weights for health states. In order to be strictly accurate we should avoid using the terms 'value' and 'utility' as if they are entirely

interchangeable. It can be said (Richardson 1994; Drummond, Sculpher et al. 2005) that 'value' is the result of decision making in a risk-free environment, and 'utility' is revealed under risk. Only in the standard gamble method is the question framed under uncertainty and so only the standard gamble can be thought of as giving utility, as based on the axioms of von Neumann-Morgenstern utility theory. To clarify, all three methods measure preference, but only the standard gamble measures utilities. Nevertheless, in the literature it is found that utility is referred to as being used to measure health regardless of what valuation method is actually used. Note that EQ-5D is valued using TTO and SF-6D using the SG.

A description has been given as to how CUA is typically operationalised, and it has been noted how the objective of the decision maker's approach is the maximization of health. It is this choice of evaluative space which leads to two concerns about the appropriateness of CUA. A further concern stems from the adoption of maximisation as the rule for decision making. The combination of using health alone as the object of value and applying societal values leads to a fourth concern which is outlined in Section 3.5.1.

One implication of restricting the evaluative space to health alone in CUA is that health interventions cannot be compared to interventions and policies which are not directly targeted at improving health, but which may be closely related (such as social services) and which are still paid for from the 'public purse'. This lack of comparability exists despite the fact that recent policy in the UK has attempted to encourage closer partnership between government departments; there is a particularly strong desire to more closely integrate health and social care services (Glendinning, Dowling et al. 2005). In light of this, economists have begun to question whether an approach to economic evaluation which artificially distinguishes between health and wider quality of life is appropriate (Grewal, Lewis et al. 2006; Coast, Flynn et al. 2008).

Despite being administered and financed by many different areas of government, one would hope that all of our public services are provided with the ultimate goal of

improving the quality of our lives⁴ (as citizens). This may involve reducing the likelihood of us becoming victims of crime, improving road safety, providing meals on wheels, a new hip, or maintaining a clean environment. All of these interventions involve the use of scarce resources and there at least comes a point when policy makers cannot increase the provision of one service without diverting resources away from others. If we are to compare the relative benefit of these diverse interventions then we need an evaluative space which is much broader than just health.

Consider the example of chronic pain, a condition which is generally not well understood in scientific terms and in which the underlying cause is often not cured by medical interventions which instead are targeted at reducing the symptoms, often at the cost of severe side-effects. We have seen that chronic pain affects most, if not all, conceivable aspects of a sufferer's quality of life. If there is limited technical ability to reduce suffering then presumably the health improvement resulting from any treatment will be small. Regardless of whether the health intervention is judged to be worth doing, there may be many policies from outside of the health sector which can have an equally or even more significant positive impact on improving quality of life. Examples include: the management by local councils of the 'Blue Badge' scheme providing a number of parking concessions for people with severe mobility problems (www.direct.gov.uk); adaptations to the home which are funded by social services, such as the attaching of handrails; regulation or assistance ensuring that those with long-term health problems are able to remain in work for as long as reasonably possible.

While it is accepted that the budget for health care (for the NHS) is fixed by central government, and it is expected that the health budget be spent on health care, this is not to say that we should adopt such a narrow focus as health to assess the outcome of health interventions. An individual patient may find it difficult to differentiate between health and social service needs (Grewal, Lewis et al. 2006) and just as there are many other factors that affect health (income, education, diet, environment) most illnesses and health interventions will affect some element of

⁴ Not necessarily defined in terms of Capability.

wider quality of life. We therefore risk underestimating the impact of illness and possibly treatment by just maximizing health (Ryan and Shackley 1995). We also risk relying too heavily on health care as a solution to every problem, when support from other agencies may be most appropriate. One review is reported by Ludbrook and Cohen (2003) and identifies 34 evidence-based measures to reduce morbidity and mortality from heart disease and stroke, only eight of which involve the use of the formal health care sector.

The very decision to use health as the maximand relies on a value judgement which may be difficult to justify. Birch and Donaldson (2003) point out that it is because “individuals are observed to prefer more health to less *ceteris paribus*... that health is [chosen as] the appropriate maximand to be used in comparing resource allocations” (p1123). It is further stated that “the normative position adopted is that the characteristic ‘health’ matters only in so far as it affects the individual’s (or patient’s) utility” (Birch and Donaldson 2003, p1124). But many factors are assumed to make up an individual’s utility set and are not included in the evaluative space of CUA. It is not difficult to consider circumstances where an individual is prepared to trade-off some level of health in order to achieve other goals or just achieve happiness in other ways. Smokers in today’s society are well aware of the negative impact of smoking on their health and although it can be argued that smoking is addictive they have still made a decision to start. It is a similar story with ‘binge drinking’. Some people choose to play sports which involve a significant risk of injury, and others are prepared to work in risky occupations in return for a high wage rate. It would appear much more appropriate to consider overall quality of life and at least seek to measure everything that is important to individuals rather than to focus on one aspect of quality of life, albeit health can be interpreted quite broadly and is unquestionably of significant importance.

Finally, the use of maximisation as the decision rule gives no allowance for considerations of equity. Recall that with welfarist CBA the Kaldor-Hicks criterion states that one state of the world can only be deemed better than another if those who are better off have gains that are sufficiently large that from these gains there is the potential for them to fully compensate any individuals who are worse off. Although in reality compensation does not have to be paid, there is scope to

consider issues of equity and equity can be considered separately from efficiency. So, those who will experience a reduction in health as a result of a particular policy (for example allowing some form of air pollution) can be compensated with some amount of money to make them indifferent between the state of the world in which they experienced clean air and that in which there is pollution. Health is just one determinant of utility and with their compensation they can increase their utility through other means.

When we move to a situation in which health is the maximand then we find ourselves in a position where compensation or any form of redistribution is impossible (Coast 2009). The production of health cannot be separated from allocation, but rather allocation happens simultaneously with production. By endorsing health maximization what is being endorsed is the “ethical position that the total sum of health produced within the health care system is what matters, no matter how that health is distributed” (Ibid. p789).

Finally, when operationalised according to the NICE guidelines, it isn't just the evaluative space which makes CUA distinctly non-welfarist; a second major departure from welfarism is the perspective from which the relative desirability of states of the world are judged. The NICE Guidance states that the weighting for QALYs (NICE 2008):

...usually comprises two elements: the description of changes in HRQL itself and a valuation of that description of HRQL. Information on changes in HRQL as a result of treatment should be reported directly by patients... The valuation of changes in HRQL reported by patients should be based on public preferences elicited using a choice-based method in a representative sample of the UK population.

The welfare (quality of life) of an individual when in a particular health state is then judged from a societal perspective rather than by the individual and any person in that same health state is deemed to have the same level of welfare.

Disease specific or generic multi-attribute health status classification systems are typically used to assess the health-related quality of life of the patient and the many health states defined by these instruments are presented to the public for valuation.

A range of generic health status classification systems are available, including the Short Form 6D (SF-6D), the Quality of Wellbeing scale (QWB) and the Health Utilities Index (HUI2 & HUI3). The EQ-5D is a generic instrument commonly used in CUA and is the instrument recommended by NICE. The EQ-5D contains five dimensions: mobility, self-care, usual activities, pain and discomfort, and anxiety and depression.

3.4 The Importance of the Evaluative Space

Within the Welfarist form of CBA money is used as the metric for assessing the relative desirability of any range of different states of the world and is used as a proxy for utility (which, it is assumed, is ultimately what every individual seeks to maximise). With CUA, however, utility is interpreted as a cardinal measure of the relative strength of preference for different health states.

3.4.1 Health

A number of weaknesses with CUA were outlined in the previous section which directly stem from using such a narrow evaluative space as health. To recap, these weaknesses were: (i) limited scope to compare health interventions with interventions falling under different policy areas (perhaps most relevant are social services); (ii) the fact that health is only one aspect of wider quality of life and to focus on health is to omit other useful and important information from the evaluation; (iii) the maximisation of health makes the separation of efficiency and equity impossible.

There is, however, a body of literature which suggests that CUA with QALYs should incorporate more in to the evaluative space than under welfarism and not less. In this chapter CUA has been presented as non-welfarist and it has been noted that the common application of CUA with QALYs is sometimes referred to as the decision maker's approach. A third term (or label) is frequently used within the health economics literature to refer to the use of QALYs together with any departure from the four tenets of the welfarist framework and this term is 'extra-welfarist'. This section makes a more specific distinction between what is non-welfarist and what is

extra-welfarist. What is here referred to as the 'extra-welfarist' framework is based predominantly on the writings of Culyer and Brouwer. That the distinction between what is non-welfarist and what is extra-welfarist is subtle is not disputed. The distinction is that extra-welfarism is based on a more theoretical perspective than the non-welfarist (decision-makers) approach, building on important work from outside the welfarist framework. In reality the result has effectively been the same narrow reliance, in both cases, on health as the maximand.

Culyer (1991) defines traditional welfarism as being based on two main assumptions:

1. Social welfare is a function of individual utilities, and
2. Individual utilities are a function only of goods and services consumed. (p67)

It will be noticed that these assumptions are essentially the third and fourth of the tenets of welfarist practice, as set out in Section 3.2.

It is noted by Culyer that within a strictly welfarist approach: "judgements about the superiority of one state of the world (defined by reference to utilities) over another are made irrespective of the non-utility aspects of each state" (Culyer 1991, p67). The extra-welfarist approach relaxes this "undue information restriction" on welfarism so that extra-welfare elements are also embodied in the judgement of social states (Culyer 1991, p67). In his extra-welfarist approach Culyer selects the characteristics of people as the important extra-welfare elements and he lists some examples of such characteristics of people as being whether they are: happy; out of pain; free to choose; physically mobile; honest (Culyer 1991, p67). Brouwer, Culyer, *et al.* (2008) provide further examples of the 'characteristics of people' which may be important which include: skills, wealth, handicap, social isolation and parenting competencies.

It is important to note that Culyer does not advocate replacing measures of individual utility, but instead sees the notion of characteristics of people as being an 'extra' source of information to aid and inform decision making.

Culyer's preferred replacement to welfarism is the QALY. In Culyer's view QALYs are able to explicitly include characteristics of people as a supplementary, non-utility, view of the quality of life. He argues that "the QALY/CEA approach need not be utilitarian at all" (Culyer 1990, p14), and that it is the ability of the QALY/CEA to

“exploit other descriptive characteristics of people ... that makes it decisively non-welfarist”.

Culyer’s solution to welfarism then is to include characteristics of people into the assessment of quality of life and to use utility as a means merely of ranking preferences or choices over different quality of life states (Culyer 1990). The framework to use for this is the QALY.

Numerous examples of characteristics have been offered above, but the one that is considered to be of paramount importance is health. Culyer appears ready to dispute the charge of maintaining a narrow focus on health, but proceeds only as far as extending the focus to process utilities:

... extra-welfarism... may be seen to take ‘health’ output as the maximand. The emphasis is not in principle exclusive, as extra-welfarism is not exclusive, and it seems unlikely that any extra-welfarist would assign equal weights to such factors as consumer choice, privacy, speed of service, hospital hotel services, and other factors that may be only remotely causally linked to health.

The importance of process utilities has been highlighted elsewhere by Mooney and others, but it is suggested by Dolan (2001) that they would be difficult to measure within the QALY approach.

Extra-welfarists therefore largely identify ‘health’ as the principal output of health services, in contrast to welfarism, under which goods and services would be taken as the natural units of output, although both are translated in to utility (Culyer 1991). The conclusion taken from this is that health is the appropriate maximand to be used in resource allocations. It is noted by Morris *et al.* that: “The set of techniques for economic evaluation which constitute applied extra-welfarism use health not to supplement utility but to substitute for it” (Morris, Devlin *et al.* 2007, p235). In other words what Culyer and Brouwer have set out as extra-welfarism (and what seems to have led to all non-welfarist CUA being labelled as extra-welfarist in the literature) is actually no different from the status quo of the decision-makers approach. This has led some authors to reject the term extra-welfarism and categorise such approaches simply as non-welfarist (Tsuchiya and Williams 2001), others have dismissed (as

was done in the start of this section) the distinction between non-welfarist and extra-welfarist approaches as “tenuous and possibly overdrawn” (Morris, Devlin et al. 2007).

Morris *et al.* state that (Ibid. p236):

It is tempting, given that the normative framework for extra-welfarism in health care was first set out in Culyer’s writings in the 1990s, to view extra welfarism as an *ex post* rationalisation for practical techniques of economic evaluation that had been applied since the late 1960s and were incompatible with traditional welfarist approaches.

Of course, the suitability of health as the evaluative space depends to some extent on how we define health and what it is that we are using health as a proxy for. As stated earlier, health economists tend to talk of health-related quality of life, implying some false assumption that health-related quality of life can somehow be separated from other aspects of quality of life. It has been suggested in the previous section and in the literature (Kind 1990) that to limit the focus of evaluations to just health is to ignore non-health factors which also contribute to a good quality of life and which are valued by the patient. If these non-health factors are affected by interventions but are not included in the evaluation then the interventions will be under (or over) valued.

Consider the example of chronic pain as an illustration of when it would appear particularly appropriate to include non-health outcomes in treatment decisions. Chronic pain is a complex and prolonged condition. Often there is simply no ‘one-off’ cure available and so in most cases it will never be observed that the patient is returned to full health. It was noted in chapter two that there are numerous different treatments for pain. Given the range of available treatments, it would seem appropriate for clinicians to base their choice of treatment not only on the severity of the patient’s pain but also according to which treatment ‘fits’ the requirements and lifestyle of the patient, in terms of side-effects, the treatment pathway and a consideration of co-morbidities. In general it seems reasonable to assume that the treatments providing the most effective pain relief will also be those with the most

unpleasant side-effects and so some balance or trade-off is required between achieving pain relief and the patient being able to function in terms of basic everyday life. It would seem likely that in reaching a decision the clinician would consider non-health outcomes. Given that non-health outcomes are likely to be considered at the level of individual diagnosis, it would also appear appropriate that they should be considered at the level of economic evaluation.

Given the World Health Organisation defines health in terms of social, emotional, and physical well-being, it is difficult to imagine that there *are* many non-health aspects of quality of life. We are not helped by the lack of any consensus on what the definitions of quality of life or wellbeing are and what they include (Kind 1990; Megone 1990; Dolan and Peasgood 2008), indeed they are two terms which often appear to be used to mean the same thing. One useful way to think of quality of life is as an umbrella term (Moons 2004, p276) covering a variety of concepts such as: functioning; health status; perceptions; life conditions; behaviour; happiness; lifestyle, etc. These concepts largely reflect the quality of the living conditions around an individual and are independent of how well the individual's own life goes. They may be considered to be a rather more objective indication of the individual's quality of life than the individual's reported utility, but this is an issue to be discussed later in this section.

Staying for now with the issue of health, one thing that is clear is how health is measured and defined by instruments such as the EQ-5D. It is stated by the EuroQol Group (Brooks 1996) that a generic measure should aim to capture physical, mental, and social functioning, i.e. those elements suggested by the WHO. Physical functioning is encompassed in the mobility and self-care dimensions, social functioning in the usual activities dimensions, and mental functioning can be assessed in the anxiety/depression dimension. The remaining dimension, 'pain and discomfort', is perhaps best considered to impact on social, mental and physical well-being. The instrument was originally designed to be simple, generic and used alongside other quality of life measures, while still achieving full coverage across the health spectrum. The EuroQol Group caution that "it is highly unlikely that such a simple instrument could be comprehensive in measuring health status, and we would not claim this from the EuroQol measure" (Brooks 1996, P56).

It was suggested earlier that health, as defined by the World Health Organisation as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity” (WHO 1946) is such a broad concept that it might be difficult to identify any aspects of quality of life unrelated to health. Therefore, if health is being measured comprehensively in CUA then it could reasonably be claimed that health is an adequate proxy for quality of life and the weaknesses of using health as the maximand may be found to be invalid; this is not, however, the reality.

3.4.2 Money as a Proxy for Utility

Many of the problems associated with CUA stem from adopting a very narrow evaluative space (i.e. health), although the evaluative space in CUA is broader than that adopted for CEA. The evaluative space for CBA is utility, a concept generally regarded as including or reflecting happiness, satisfaction and desire fulfilment. Money is used as a proxy for utility in CBA. Utility is obtained from the consumption of goods and services and from the outcomes of behaviour or processes. Given the definition of utility, all conceivable outcomes can be valued using CBA. But, even if it is accepted that money is an accurate proxy for utility (and this has been questioned in Section 3.3), it has been argued by some that utility does not provide us with a useful evaluative space (Nussbaum and Sen 1993; Sen 1999). Put another way, it can be questioned whether happiness and desire adequately reflect welfare (quality of life).

Sen has argued (Sen 1999) that both views of utility (as happiness or as desire fulfilment) have the characteristics of “(1) being fully grounded on the mental attitude of the person, and (2) avoiding any direct reference to the person’s own valuation exercise – the mental activity of valuing one kind of life rather than another” (P14). Sen has labelled the first characteristic ‘physical-condition neglect’ and the second ‘valuation neglect’.

By ‘physical-condition neglect’ Sen means that the “physical conditions of a person do not enter the view of well-being seen entirely in terms of happiness or desire-

fulfilment, except insofar as they are indirectly covered by the mental attitudes of happiness and desire” (Sen 1999, P14). A person who has a chronic medical condition, who is undernourished or who has no access to shelter may still be high on the scale of happiness if they have adapted to and accepted their unfavourable physical conditions. Setting out a similar argument to Sen, Culyer states that: “To compare the ill health of different individuals or groups is not the same as to compare ... their pleasure (a sick optimist may have far more pleasure from life than a well grumbler)” (Culyer 1990, p12). It is not at all obvious that the sick optimist is higher on the scale of well-being than the well grumbler.

In terms of ‘valuation neglect’, Sen argues that valuing is not the same thing as desiring and the strength of desire is influenced by considerations of realism. Considerations of ‘feasibility’ and of ‘practical possibility’ enter into what we dare to desire (Ibid. P15). If a person with a long-term illness has learnt to adapt to their limited mobility, social isolation and the adverse side-effects of their medication then they are unlikely to have overambitious desires and may take pleasure from simple every-day things. It is feasible that such an individual will have their desires fulfilled more frequently and report a greater degree of happiness than a healthier individual with more ambitious desires.

Research in the field of health care provides much evidence that many patient groups adapt to their illness or disability. It is noted (Groot 2000; Damschroder, Zikmund-Fisher et al. 2005) that patients with chronic health conditions often rate their quality of life significantly higher than non-patients rate the quality of life associated with the patients’ health state. The explanation given for this by both sets of authors is that patients adapt to their situation, largely as a means of coping. When people, for example, are asked to imagine life on dialysis, healthy people predict being unhappy the majority of the time, and yet most dialysis patients report being happy the majority of the time (Damschroder, Zikmund-Fisher et al. 2005). So what will be the implications of adaptation? If all else is equal, if adaptation to one condition is more complete than to another, then patients with the former will be given less priority than those with the latter in the competition for resources to improve quality of life (Dolan and Kahneman 2008). Is this the ‘wrong’ result? Dolan

and Kahneman suggest not, so long as it is the patient who is valuing their own well-being.

As we have seen, evidence suggests that individuals adapt to *some* unfavourable aspects or circumstances in their life, and this adaptation is frequently referred to as 'response shift' (Sprangers and Schwartz 1999). This response shift has been found in numerous groups, including paraplegics, some transplant patients and type I diabetes patients, as well as patients who are on dialysis (Groot 2000; Damschroder, Zikmund-Fisher et al. 2005; Dolan and Kahneman 2008). It is also widely accepted that people adapt to increases in their personal wealth and that in the long-term increases in wealth do not increase happiness (Layard 2006). But, there are some illnesses or life events that people do not react to; unemployment is one such state and evidence suggests that people actually become increasingly sensitive to pain and to noise pollution/disturbance (Dolan and Kahneman 2008). Dolan and Kahneman argue that it is no fairer or more justifiable to decide to give more weight to conditions that people *do* adapt to than to instead target resources on helping those with unfavourable circumstances that they *cannot* adapt to. Dolan and Kahneman conclude that: "the problems associated with allowing for adapted preferences have been overstated, especially when deciding how to allocate resources once the budget for health care has been determined" (Ibid. p216).

There is, however, a problem when non-patients are asked to imagine being in a particular health state, in that they tend to focus on what would be the 'initial shock' and on the factors which differentiate that health state from their current health. Dolan and Kahneman (2008) cite a suggestion from the psychology literature⁵ that when predicting the impact of ill health those not directly affected will underestimate the power of our 'psychological immune system'.

When Sen discusses adaptation it appears to be the likelihood of individuals demanding health care that he is concerned with, rather than the reported well-being of patients who have a diagnosis and are already receiving treatment. That patients are able to adapt is likely to be a reflection on the quality of treatment they receive,

⁵ Wilson, T. D. & Gilbert, D. (2003) 'Affective Forecasting', *Advances in Experimental Social Psychology*, vol. 35, pp. 345-411

the support offered by health care professionals, friends and family and some common human belief that life despite its trials is worth living. Following on from his discussion of physical-condition and valuation neglect in the earlier literature, Sen (2002) suggests, specifically in the context of health, that:

[T]he patient's internal assessment may be seriously limited by his or her social experience. To take an extreme case, a person brought up in a community with a great many diseases and few medical facilities may be inclined to take certain symptoms as "normal" when they are clinically preventable.

We are asked to consider the case of two states in India and told that the state of Kerala has the highest rate of literacy and longevity in India, but also has the highest rate of self-reported morbidity among all Indian states (Sen 2002). At the other extreme is Bihar, a state with low longevity, "woeful medical and educational facilities" and which has the lowest rate of self reported morbidity in India. The explanation which Sen offers is that people in states that provide more education and better health facilities are in a better position to self diagnose than people in states where there is less awareness of treatable conditions (Ibid.).

There may then be good reasons to doubt whether an individual's own utility will be closely correlated to some external and more objective assessment of wellbeing, based on the individual's achievement on a range of dimensions such as health, living conditions, education, friendship and support, and wealth. It was suggested in section 3.4.1 that quality of life be thought of as an umbrella term covering a number of concepts many of which, such as health status, life conditions and education, can be measured more objectively than by relying on utility.

3.4.3 Benefits Explicitly Valued in WTP Studies

Recall that the major advantage of CBA over CUA is the potential for a much more comprehensive consideration of outcomes. Olsen and Smith (2001) assess the range and detail of benefits described in empirical studies in order to gain an understanding of the extent to which this 'comparative advantage' has been exhausted in practice. The authors report a "huge mismatch between the theoretical

glory of WTP and the usefulness for public health policy of the majority of surveys which have applied this method” (Olsen and Smith 2001). Again, this is an issue relating to study design rather than any kind of fundamental weakness with the approach.

Olsen and Smith assess 71 studies according to the description given to respondents of the programmes for which they were asked to state their WTP. It is suggested that the descriptions should include health outcomes (incorporating health state improvement, duration and probability) and beyond health outcomes (referred to elsewhere as process utilities). Only seven studies in Olsen and Smith’s review had descriptions which contained all three dimensions of health outcome. Only four surveys had scenario descriptions which were sufficiently comprehensive to facilitate the estimation of QALYs (Ibid. p42). Only 18 of the 39 studies which included beyond health outcomes (process utility) gave a comprehensive or understandable description of these outcomes, as judged by Olsen and Smith (2001, p43). The authors note that the majority of studies were concerned with the WTP for a health care *product* rather than with the benefits (for health or quality of life) produced by the product (Ibid. p44).

3.4.4 Experienced Utility

A new approach is attracting attention within health economics which is distinctly different from either CUA or CBA and which has its theoretical roots more firmly planted in the utilitarian tradition. This new approach is experienced utility, or “subjective evaluation” (Dolan and Peasgood 2008). This essentially involves asking people how satisfied they feel with their lives overall. Here, there is a rejection of neoclassical welfare economics, which assumes that decision utility can be inferred from the preferences that agents reveal in their choices. Experienced Utility has also attracted interest recently following concerns regarding the legitimacy of the values for health states elicited from members of the public, and here we have a direct link with the discussion concerning adaptation in the previous sub-section.

An approach which has been advocated by Dolan and Kahneman (2008) is the day reconstruction method (DRM), in which respondents are asked to divide the previous

day into a number of episodes and then to rate different elements of affect during those activities on a cardinal scale. It is suggested that if sufficient data is gathered then researchers may be able to identify the utility associated with specific activities. It is also suggested that by gathering longitudinal data on moment-by-moment utility that some 'background' influences or distortions, such as relative wealth will be minimised and that the situation in which a respondent is drawn to focus on just one aspect of their life in response to questioning will be avoided. A source of data which is being explored at present is the British Household Panel Survey (Dolan 2008c).

3.5 Perspective

3.5.1 Cost-Utility Analysis

If CUA is operationalised according to the NICE guidance, then the individual is no longer the judge of their own utility, but instead some generic societal value is attached to their health-related quality of life. This societal value is the average of the utilities elicited from the general public. It should be clear then that we need a consistent method by which patients can report their health-related quality of life and with which we can describe a health state to the respondents completing the valuation task. The health states presented to the respondent for the valuation task should provide a clear representation and summary of what effect the disease and/or treatment will have: physically, psychologically, emotionally and socially (Smith and Dobson 1993, p349). This description is the sole basis for the respondent's subjective valuation of the perceived impact that the health state would have on their quality of life. It has been suggested in previous sections that by using the EQ-5D, a short generic instrument containing just five dimension, the descriptions of the health states presented to the public for valuation cannot be expected even to represent a comprehensive description of the impact of an illness or intervention on health. In other words, given that respondents have no direct experience of the health state they are valuing they instead rely on a description of the health state given to them by the researcher, but this description is often inadequate.

The combined effect of eliciting values from non-patients and providing these respondents with only a partial understanding of the effects that a health state has

on quality of life can potentially create a situation in which the opportunity cost of an individual patient is undermined. When individuals make choices they may sometimes have well considered reasons for not preferring the healthiest options; this will be the case due to opportunity costs associated with the different options. Birch and Donaldson argue that under an extra-welfarist approach no consideration is made for the individual's foregone alternatives. "In this way, individuals' own choices, even where these choices are an accurate representation of the individuals' utility-maximising positions..., can be violated in the interests of maximising health" (Birch & Donaldson 2003, p1124). The authors regard this scope to override individuals' considerations of opportunity cost as a rejection of the basic foundations of the economics discipline. It is stated later in the paper that an approach in which health is used as the maximand "represents multiple sources of 'harm' to the individual as the utility of non-health commodities is ignored and the individual's right to choose is undermined" (Birch & Donaldson 2003, p1131).

The issue is not so much caused by using societal values to assess the wellbeing of an individual but more because the members of society are valuing one single and potentially narrow aspect of that individual's wellbeing. Providing respondents with a more comprehensive description of the outcomes associated with an illness or intervention may help to limit the scope for decisions made by a policy-maker to override the individual's considerations of opportunity cost.

If, on the other hand, patients report their own utility while in a particular health state then the utility they report may be susceptible to adaptation and coping mechanisms. But at least it will be based on a consideration of exactly what it is like to experience a particular health state or the effects of a health intervention. While that valuation may follow adaptation it is a true reflection of the experience and the adaptation itself may be a positive and useful process (or even an intervention, such as Cognitive Behavioural Therapy). The valuation may also depend, however, on whether the individual has a cheerful disposition or is prone to take a negative view of the world and it may be considered that this influence distorts the usefulness of the valuation as a true reflection of their wellbeing. It may also be the case, as discussed above, that a person conditioned by their unfavourable circumstances and/or lack of education may simply tolerate poor health whereas a person who has greater

knowledge and/or who enjoys a more affluent lifestyle may have a greater expectation of how healthy they should be and have a greater demand for health care. This distortion has led some to seek an alternative evaluative space entirely (Sen 1983; Sen 1987; Nussbaum and Sen 1993; Sen 1999; Sen 1999a; Alkire 2005; Robeyns 2006), that of capability.

3.5.2 Cost-Benefit Analysis

There has already been a fairly comprehensive discussion in Section 3.2 about the importance of perspective within CBA. Recall that consumers may not act in the same way as citizens and even though health is mostly a private good, it is one which is provided by a national health service in the UK. The way in which people respond if they believe they are being asked to contribute as taxpayers may be troublesome as there are various incentives such as altruism, consideration for relative wealth and even free-riding. It has been noted however that the use of average private willingness to pay values across diverse groups to inform policy may be helpful in reducing problems with equity arising from ability to pay and may avoid the problems encountered when asking people to respond as taxpayers.

One issue not yet discussed is that of the correct amount of information to present to non-patient respondents. If respondents are being asked for their willingness to pay for what is to them hypothetical treatments then the same issue arises as with CUA as to how this scenario is presented to the respondent and in how much detail. Recall that Olsen and Smith assessed WTP studies in terms of the description of the programmes to be valued by respondents. The assessment of descriptions in terms of health and beyond health outcomes has already been discussed. The authors suggest that if WTP were to exhaust its potential regarding the valuation of a wide range of valuable characteristics, these characteristics need to be clearly spelt out to respondents (Olsen and Smith 2001, p44). The authors further suggest that to present this amount of information to the respondent will require a separate scenario description (separate from and prior to the actual WTP question). In only 31 of the 71 studies reviewed were respondents given a separate scenario description. It should be acknowledged that the majority of studies involved patients, who may be expected to have had more knowledge of the health treatment *ceteris paribus* than

members of the public would have had, although this doesn't alter the fact that in the majority of cases the WTP appears to have been for the health product rather than the benefits produced by the health product.

3.6 Discussion

If a useful approach to economic evaluation should allow us to compare the efficiency of different interventions funded from the same health care budget, and if it is accepted that any health state or health care intervention is likely to have a range of effects which are important to the patient then CEA can be dismissed. CUA is the most commonly used approach in the UK and is recommended as the approach to be used in all technology appraisals submitted to NICE, although there are three weaknesses with CUA which have been set out in this chapter. First, with CUA there is limited scope to compare health interventions with interventions falling under different policy areas (for example social services). The evaluative space is the second weakness; health is only one aspect of wider quality of life and to focus on health is to omit other useful and important information from the evaluation. Third, the maximisation of health makes the separation of efficiency and equity impossible.

When operationalised in accordance with the NICE guidance it is also the case that non-patients are chosen to complete the valuation exercise. What these respondents value is a fairly limited description of a health state, when health is just one aspect of the patient's quality of life. This explains why respondents predict significantly lower levels of happiness for patients than the happiness reported by the patients themselves. It also makes it much more likely that the decision making process at a societal level will override patients' own considerations of opportunity cost.

In CBA there is the potential for a much broader evaluative space, although there are many methodological challenges to overcome when operationalising the approach, which may explain why in practice this conceptual advantage has not been realised.

In short, there is a clear justification for continuing research and this research may involve improving study design for WTP, broadening the existing QALY for use in

CUA, or exploring new alternatives. One such alternative, the possibility of measuring experienced utility, has briefly been introduced. A second alternative, the capability approach of Amartya Sen, is also receiving increasing attention within the discipline of health economics (Verkerk, Busschbach et al. 2001; Cookson 2005; Coast, Flynn et al. 2006; Grewal, Lewis et al. 2006; Coast, Smith et al. 2008; Kinghorn and Smith 2008; Lorgelly, Lorimer et al. 2008; Coast, Smith et al. 2008a; Coast, Smith et al. 2008c; Coast 2009). Culyer and Brouwer drew extensively on the writing of Amartya Sen in justifying the need for extra-welfarism, although extra-welfarism is distinctly different from capability and a purer form of the approach may provide a useful means of describing and assessing quality of life in a way which is more objective than the use of utility. Future chapters will set out the capability approach in greater detail and explore whether and how the approach can be operationalised in order to assess quality of life in patients with chronic pain.

CHAPTER 4: Systematic Review: Outcome Measures Used in Economic Evaluations Relating to Chronic Pain

4.1 Introduction

In chapter two a number of instruments for assessing pain, health and health-related quality of life were introduced. It was found that there is wide variability in terms of specificity (to a particular condition), scope (from assessment of a single aspect of health through to health-related quality of life), and use (either clinical or economic evaluation). Chapter three introduced and discussed the different types of economic evaluation, along with the theory underpinning, and the general issues surrounding, each.

It has been suggested in chapters two and three that where the only outcome considered in an economic evaluation is some narrow measure of health, then some important information about changes in the lives of patients will be excluded. Aspects of daily life and experiences likely to be affected by chronic pain were identified in Chapter two as including (but not necessarily limited to):

- Pain itself (intensity and frequency)
- Personal identity and legitimisation of the condition
- Activities of daily living
- Independence
- Employment
- Sleep disturbance
- Relationships
- Social issues and loneliness
- Psychological and emotional distress
- Enjoyment of life.

It may be, however, that some of these dimensions are included in economic evaluations in other ways. For example, if loss of income is included as an indirect cost then to some extent employment *will* be included in the evaluation; it will just appear on the other side of the equation. Including issues relating to employment as an outcome may then lead to some degree of double counting. This may especially

be the case with willingness to pay. If a patient considers the loss of employment, due to ill health, to be a significant impact on their quality of life and is asked to value the overall impact in monetary terms then it is possible they will consider and include loss of earnings; even if advised not to.

A possible concern relating to instruments which seek to assess quality of life rather than health, and one raised at forums such as the Health Economists' Study Group and by John Brazier at the International Health Economics Association, is the extent to which wider measures of quality of life will contain attributes which will actually be affected by changes in *health*. It may be that broader and more permanent aspects of life dominate to such an extent that they mask an improvement in health which would be significant if considered in isolation. Or that it takes longer for a change in health to have a 'knock-on' effect on broader aspects of quality of life. If this is the case then changes in the quality of a patient's life may not be detected during the time periods over which economic evaluations are typically conducted.

Although health is the defining goal of a health care system, in itself, in terms of the presence or absence of disease and symptoms, it may not be the most important aspect of quality of life for patients. As suggested by Dolan and Kahneman (2008), there may be some symptoms (or alternatively events in life) which patients 'easily' adapt to. It may then be that very narrow measures of health detect changes whereas wider quality of life measures do not, and the latter case reflects the true preferences or satisfaction of the patient. Here, the change in symptoms would simply not be of significant importance to the patient. Recall, however, that pain was listed by Dolan and Kahneman as one of the symptoms which patients do *not* adapt to. If this is the case, it would be reasonable to assume that whenever severe pain is present, quality of life will be reduced.

The main focus of the review outlined in this chapter is to explore the outcome measures used in economic evaluations relating to chronic pain (i.e. measurement of benefit). The reasons stated for using one particular type of evaluation will also be of interest. It has been noted previously how CUA is the approach recommended for use in the UK by NICE, but the search is not restricted to evaluations conducted in the UK. It may be that certain types of evaluation are either used or avoided due to

reasons discussed (other than the NICE guidance), or not yet considered, so far in this thesis.

This chapter will look at specific examples of actual economic evaluations relating to chronic pain. The intention here will be to answer the following questions:

1. What types of economic evaluation have been conducted and what outcome measures have been used?
2. On what basis have researchers chosen and justified the use of a particular outcome measure?
3. Does it appear as though some attempt has been made to detect broader changes in quality of life, or have studies simply assessed pain and disability? To what extent have those dimensions identified in chapter two been included?
4. If more than one outcome measure has been used, is there a significant difference in the level of responsiveness and/or differences in the apparent desirability of an intervention as indicated by the different measures?
5. Over what timescales have results been collected? Do some instruments begin to detect changes in outcome with a shorter delay (lag) than others?

The chapter will proceed as follows: Section 4.2 will outline the search methods used, along with the inclusion and exclusion criteria used to select a final set of studies for discussion; Section 4.3 will provide a brief summary of results and Section 4.4 will contain a more detailed discussion of the studies. Section 4.5 concludes.

4.2 Search Methods

A search was conducted to find articles which outline some form of economic evaluation relating to an intervention for the relief of chronic pain. The purpose of the search was to try and answer the questions listed in the introduction to this chapter.

A search for economic evaluations was conducted using six on-line literature databases, listed below:

- **Academic Search Elite (EbscoHost):** provides full text for journals covering the social sciences, general science and much more. Articles can be retrieved which date from 1985 to the present year.
- **EconLit (Ebsco):** is an Economics database that includes citations and abstracts for journal articles, books, conference proceedings, etc., since 1969.
- **MEDLINE (Ovid):** is a database covering the fields of medicine, nursing, the healthcare system and the biomedical sciences. Time-span is generally 1950 – present year. MEDLINE is the primary component of PubMed, although it can be searched via a separate ‘gateway’.
(www.nlm.nih.gov/pubs/factsheets/medline.html)
- **PubMed:** is a service of the National Library of Medicine (US) which provides access to over 12 million citations dating back to the mid-1960’s. Key areas covered are: medicine, biology, nursing, drug, social work.
- **Science Direct (via Scirus):** is an information source for scientific, technical and medical research.
- **Web of Science/ Web of Knowledge:** is a database providing citations for journal articles, books and other media. Includes three databases spanning all academic subjects and disciplines: Science Citation Index; Social Sciences Citation Index; Arts & Humanities Citation Index. Coverage is from 1945 onwards.

Because the emphasis of the search was on exploring what work has been reported on the topic of chronic pain, as opposed to testing the hypothesis that a new economic evaluation is worth doing, there was an effort to retrieve a large number of studies. Had the intention been to test the hypothesis that it would be worthwhile conducting an economic evaluation of a particular intervention for the relief of chronic pain, a more stringent inclusion criteria would have been chosen, including some check on the quality of studies selected.

A list of search terms and a record of search results are displayed in Table 4.4. No results (or ‘hits’) were retrieved via the EconLit database when searched through MetaLib. Only four results were found when a direct search of EconLit was conducted, one of which was a duplicate, three of which were unsuitable and so this database is excluded from Table 4.4. No results were retrieved when MEDLINE was

searched through MetaLib, but results were found when a direct search was conducted of this database. A number of articles were found either through different databases when searched using the same search terms, or in repeated searches using different terms. After the first time an article was found, whether it was saved or not, it was recorded as a duplicate each additional time it appeared. Articles were initially assessed based on their title and abstract alone and were excluded at this stage if they had either already been retrieved using a different database or different search terms (and hence were duplicates), or if they met the exclusion criteria summarised in Table 4.1.

Table 4.1 Exclusion Criteria:

- Non-English Language
- Only acute pain (of < 6 months) or chest pain or heartburn.
- Letters to the editor
- Main focus on a chronic condition, of which pain is merely listed as a possible symptom (including cancer care)
- Brief mention of economic considerations, but no attempt at economic evaluation (i.e. no comparison of interventions in terms of costs and consequences)

Any articles not excluded at the initial search stage were read in full. It can be seen from Table 4.4, that after excluding duplicates, a total of 919 articles were found, of which 83 were saved and read in full. Eighteen articles were selected for discussion in this thesis, all of which are summarised in table 4.4. These 18 articles were selected according to the inclusion criteria found in Table 4.2. A further three articles were found through a manual search of the references of any review articles retrieved during the original search; the number of articles summarised in Table 4.4 therefore totals 21. Figure 4.1 summarises the search process. Upon reading the articles it became apparent that the same study had been found twice, as it had been published in two different journals by different first authors (Thomas, MacPherson et al. 2005; Ratcliffe, Thomas et al. 2006). This reduces the final number of studies to 20.

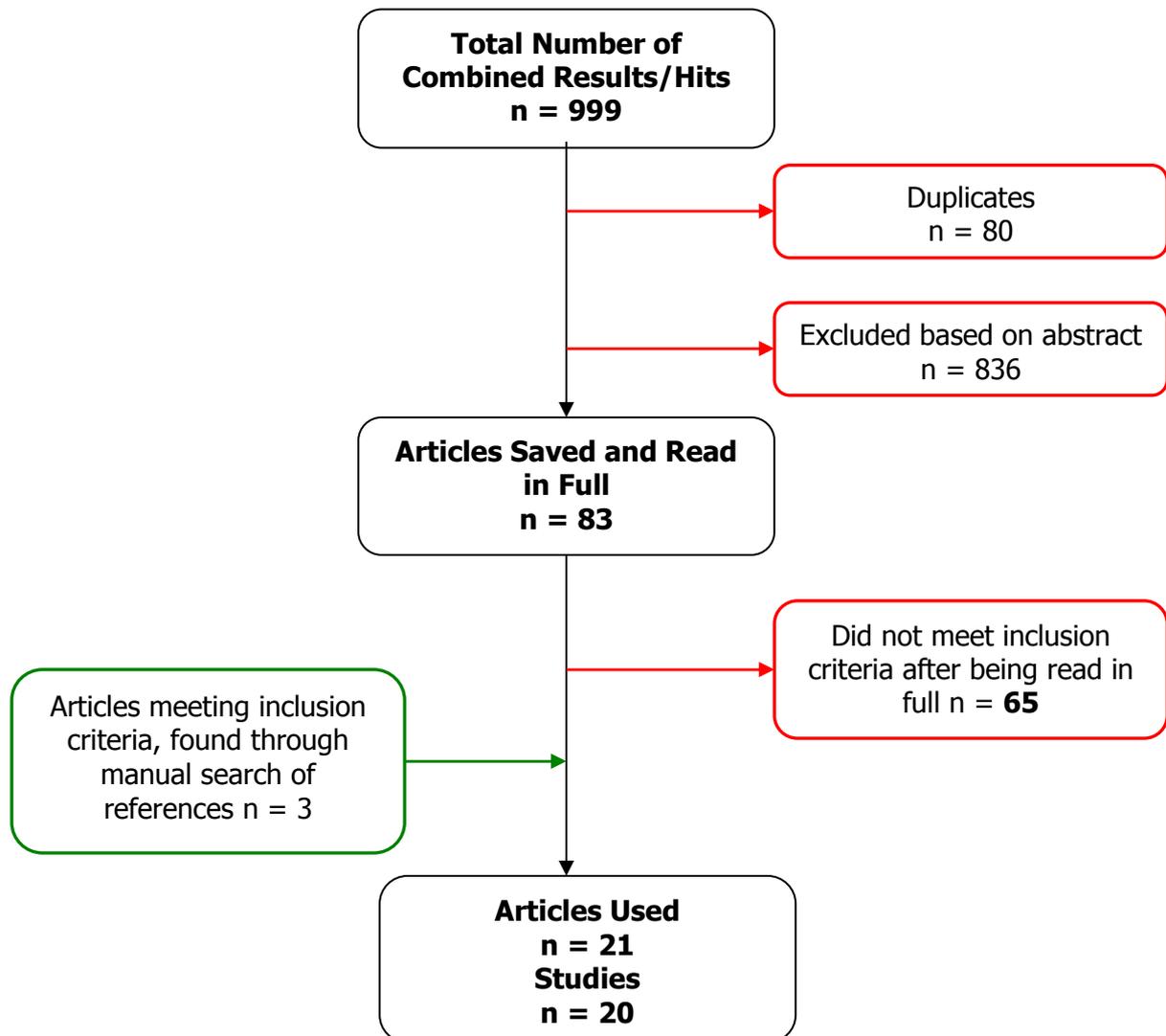
Table 4.2 Inclusion Criteria:

- English Language
- Discussion of economic costs AND outcomes
- Studies including some attempt at: cost-minimisation, cost-effectiveness, cost-benefit, or cost-utility analysis
- Or, articles reviewing: cost-minimisation, cost-effectiveness, cost-benefit, or cost-utility analyses. (These are not included in Table 4.4, but references from these papers were checked to find any new articles not retrieved through the search outlined above.)
- Reference to chronic or 'long-term' pain (or at least a sample of both chronic and acute pain patients)

Table 4.3 Search Terms and Summary of Search Results: Economic Evaluation

Date Searched	First Search Term		Second Search Term		Individual Results					Combined Results	Duplicated Results	Initially Excluded
	By	Term	By	Term	EbscoH	PubMed	MEDLINE	Science Direct	Web of Know'			
Mar 07	Title	chronic pain	Subject	economic evaluation	0	0	-	2	19	21	2	5
Mar 07	Subject	chronic pain	Subject	economic evaluation	0	0	-	4	61	65	13	41
Mar 07	Subject	chronic pain	Subject	cost-utility	0	0	-	3	28	31	13	16
Mar 07	Subject	pain	Subject	cost-utility	0	0	-	16	179	195	11	170
Mar 07	Subject	chronic pain	Subject	cost-effectiveness	0	0	-	34	200	234	14	207
Mar 07	Subject	chronic pain	Subject	cost minimisation	0	0	-	0	0	0	0	0
Mar 07	Subject	pain	Subject	cost minimisation	0	0	-	0	9	9	0	8
Mar 07	Subject	long-term pain	Subject	economic evaluation	0	0	-	35	24	59	2	56
Apr 07	Subject	chronic pain	Title	cost-benefit	0	0	-	3	3	6	0	6
Apr 07	Title	chronic pain	Subject	cost-benefit	0	42	-	3	8	53	8	39
May 07	Title	cost-effectiv*	Subject	chronic pain	1	0	-	2	57	60	7	50
May 07	Subject	economic evaluation	Title	fibromyalgia	0	0	-	0	2	2	0	1
May 07	Subject	cost-utility	Title	fibromyalgia	0	0	-	0	0	0	0	0
May 07	subject	cost-effectiveness	Title	fibromyalgia	0	0	-	0	9	9	0	8
May 07	Subject	economic evaluation	Title	painful diabetic neuropathy	0	0	-	0	0	0	0	0
May 07	Subject	cost-effectiveness	Title	painful diabetic neuropathy	0	0	-	0	1	0	0	0
May 07	Subject	Cost-utility	Title	painful diabetic neuropathy	0	0	-	0	0	0	0	0
May 07	Subject	economic evaluation	Title	neck pain	0	0	-	1	3	4	2	1
May 07	Subject	cost effectiveness	Title	neck pain	0	0	-	2	15	17	4	10
May 07	Subject	cost-utility	Title	neck pain	0	0	-	0	1	1	1	0
Feb 08	Search Term	Pain	Sub-heading	Economics	-	-	233	-	-	233	3	216
Total										999	-80	-836

Figure 4.1 Summary of Literature Search for Articles Discussing/Detailing Economic Evaluations of Interventions for the Relief of Chronic Pain:



4.3 Search Results

It is not clear that all of the 20 studies selected for review report studies involving respondents with ‘chronic pain’ as defined in this thesis (pain persisting for at least six months). Not all of the papers summarised in Tables 4.4 and 4.5 below include a definition of chronic pain and few give an indication as to the duration of the pain experienced by respondents in the studies. Given this lack of information, studies were only included if the pain was described by the authors as being chronic.

As each of the articles in Table 4.4 was read they were labelled according to the type of economic evaluation described (cost-utility, cost effectiveness, cost-benefit, or cost-minimisation). Often the authors of the papers simply referred to any type of evaluation as cost-effectiveness, whereas in summarising the articles for Table 4.4 an attempt has been made to classify them more accurately.

Once again, the authors did not all use consistent terminology when describing the types of cost that were included, and so an effort was made to more accurately categorise costs as the papers were read, according to the following definitions given by Jefferson *et al.* (2000):

- *Direct Costs:* borne by the health care system, community and family in directly addressing the problem.
- *Indirect Costs:* mainly productivity losses caused by the problem or disease, borne by the individual, family, society, or by the employer.

It is important to report the specific patient population in Table 4.4 as there are many diverse conditions with which chronic pain is associated and a minority of outcome measures are specific to one of these (e.g. to back pain). It may also be the case that some more generic outcome measures work better with some patient groups than with others.

The comparators being assessed largely define the economic evaluation and so are included in Table 4.4. Finally, the result (the incremental analysis) is important for two reasons. First, it is interesting to see the nature of the conclusion that it is possible to reach given the design of the study. Second, it is possible that the same comparators have been compared in different studies and yet different conclusions have been reached.

Table 4.5 provides a brief summary of each of the studies in terms of their basic design and the characteristics of participants. When reported by the author, 95% confidence intervals are given in brackets in the results column.

Table 4.4: Economic Evaluations Relating to Chronic Pain

Author (Year)	Intervention being assessed	Comparator	Type of Economic Evaluation	Patient Group	Outcome(s)	Result	Country
Riverero-Arias, O. et al (2005)	Surgical stabilisation of the spine	Intensive rehabilitation	Cost-utility	Chronic low back pain	EQ-5D	ICER of immediate surgery estimated as £48,588 per QALY gained (-£279,883 to £372,406)	UK
Rivero-Arias, O. et al. (2006)	Routine physiotherapy treatment	Advice from a physiotherapist	Cost-utility	Subacute & chronic low back pain	EQ-5D (ODI)	Estimated ICER of £3010 per QALY	UK
Ratcliffe, J. et al. (2006) & Thomas, K. J., et al (2005)	Acupuncture	usual care only	Cost-utility	non-specific low back pain of 4-52 weeks duration	SF-6D (converted from patient responses to SF-36)	Mean ICER for acupuncture of £4241 per QALY gained (£182 to £27,899)	UK
Schweikert, B. et al. (2006)	Addition of cognitive behavioural treatment (CBT)	Standard 3-week inpatient rehabilitation	Cost-utility	Non-specific low back pain of at least 6 months	EQ-5D	ICER of €-126,731 per QALY gained	Germany
Willich, S. et al. (2006)	Additional acupuncture treatment	Routine care alone	Cost-utility	Chronic neck pain	SF-6D (converted from patient responses to SF-36)	ICER was €12,469 per QALY gained	Germany
Greiner, W. et al. (2006)	Transdermal fentanyl (TDF)	3 other long-acting opioids (TBD; CO; SMM)	Cost-utility	Moderate to severe non-malignant chronic pain	Utility values reported in the literature for controlled & uncontrolled pain and for similar symptoms to the side effects considered	TDF provides patient with 0.539 QALYs vs. 0.537 for TBD, 0.502 for CO, and 0.499 for SRM. ICER is €1,625.65 for TDF vs. SRM (-€1,041 to €17,286)	Germany
Witt, C. et al (2006)	Acupuncture in addition to routine care	No-acupuncture control group	Cost-utility	Chronic low back pain	SF-36, transformed to SF-6D	ICER of €10,526 per QALY gained (probability that acupuncture is cost-effective is close to 100%)	Germany

Table 4.4 continued

Author (Year)	Intervention	Comparator	Economic Evaluation	Patient Group	Outcomes(s)	Result	Country
Tarride, J. et al. (2006)	Pregabalin	Gabapentin	Cost-utility	Chronic neuropathic pain	EQ-5D	Lower cost & greater effectiveness of treatment with pregabalin	Canada
Neighbors, D. et al. (2001)	Fentanyl Transdermal System	Two long-acting oral opioids	Cost-utility	General moderate to severe chronic pain	Estimated utility (based on disutility of severe pain from shingles derived using the SG – reported elsewhere by other authors)	Compared to controlled-release morphine, ICER of \$20,709 per QALY. Compared to controlled-release oxycodone, ICER of \$5,273 per QALY	USA
Spiegel, B. et al. (2003)	Rofecoxib & celecoxib	nonselective nonsteroidal anti-inflammatory drugs (NSAIDs)	Cost-utility	Osteoarthritis or rheumatoid arthritis patients with moderate to severe pain	Use of 'validated Utilities developed by previous investigators'	Use of coxibs instead of NSAIDs may cost an additional \$275,809 per yr to gain 1 extra QALY	USA
Taylor, R. J. & Taylor, R. S. (2005)	Spinal Cord Stimulation	Nonsurgical, conventional medical management	Cost-utility	Patients with failed back surgery syndrome (FBSS)	Utility values identified through a search of the literature and imputed based on methods stipulated by other authors.	Incremental cost-effectiveness ratio for spinal cord stimulation of €45,819 Euros per QALY (best case = €30,370; worst case = €63,511)	
Kemler, M. & Furnee, C. (2002)	Spinal Cord Stimulation (SCS) with Physical Therapy (PT)	Physical therapy alone	Cost-Effectiveness & Cost-Utility	Patients with chronic reflex sympathetic dystrophy	Pain relief by VAS & health-related quality of life by EQ-5D	SCS & PT both more effective and less expensive over a lifetime than PT. Incremental cost per QALY gained = 9,350 Euro per QALY for worst-case scenario.	Netherlands

Table 4.4 continued

Author (Year)	Intervention	Comparator	Economic Evaluation	Patient Group	Outcomes(s)	Result	Country
Thompson, M. et al (2005)	Oral Rizatriptan	Other oral triptans	Cost-Effectiveness & Cost-utility	Patients with moderate to severe migraine	Number of attacks aborted & Utility values for migraine, based on a study using QWB scale	Incremental cost per attack aborted of \$Can 49.82 and per QALY gained of \$Can 31, 845. From societal perspective, rizatriptan was more effective and produced cost savings.	Canada
Niemisto, L. et al. (2005)	Combination of interventions (inc. physician consultation)	Physician consultation alone	Cost-effectiveness	Patients with chronic low back pain	Pain intensity on VAS (15D – quality of life; ODI – disability)	A 1-point improvement in combination grp, compared to consultation grp, on VAS costs \$512 (\$77 to \$949)	Finland
Lissovoy, G. et al. (1997)	Intrathecal Morphine Therapy (IMT)	Alternative therapy	Cost-effectiveness	Failed back surgery resulting in chronic pain	Years of good to excellent Pain Relief (average of the upper & lower estimates identified from the literature).	Projected cost per year of pain relief for IMT versus conventional pain management over 60 months and at base case is -\$624 (best case = -\$7212)	USA
Haas, M. (2005)	Treatment by doctors of chiropractic	Treatment by primary-care medical doctors	Cost-effectiveness	Chronic and acute low back pain	Pain intensity on VAS (ODI; SF-12)	For chronic patients at 3 months, \$5 additional cost for a 10.5-point reduction in pain.	USA
Thomas, K. et al (2005)	Exercise Therapy	Monthly telephone support; exercise & telephone support; no intervention	Cost-effectiveness	Knee Pain	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)	Exercise intervention cost effective at a willingness to pay of £8,000 for a 50% improvement in knee pain.	UK

Table 4.4 continued

Author (Year)	Intervention	Comparator	Economic Evaluation	Patient Group	Outcomes(s)	Result	Country
Goossens, M. et al (1998)	Operant programme with cognitive programme & relaxation	Waiting List; Operant treatment with group discussion treatment	Cost-minimisation	Chronic low back pain	Adapted version of McMaster Utility Measurement Questionnaire used to record health states. Utility values elicited from patients using rating scale and SG.	Operant treatment is more effective in improving utilities than no treatment. Adding a cognitive & relaxation programme of approx. US\$1251 did not result in any additional benefit.	Netherlands
Kumar, K. et al. (2002)	Intrathecal drug therapy	Conventional pain therapy	Cost-minimisation	Chronic low back pain caused by failed back syndrome	(ODI – for disability; pain relief on VAS)	Cumulative cost per patient for 5-yr period of IDT \$29,410. \$38,000 for conventional therapy	Canada
Goossens, M. et al. (1996)	6 week educational/ cognitive intervention (ECO)	Educational discussion intervention (EDI) and waiting list (WLC)	Cost-Minimisation	Fibromyalgia	Maastricht Utility Measurement Questionnaire. Utility measurement using rating Scale & Standard Gamble	No significant differences in utilities between two groups, so authors simply report that direct medical costs are \$1,623US higher for ECO.	Netherlands

Table 4.5: Study Characteristics

Author (Year)	Sample Size	Study Design	Sensitivity Analysis?	Time-period (inc. frequency of data collection)	Average time since onset of pain	Age Range (Years)
Riverero-Arias, O. et al (2005)	349	Randomised Controlled Trial	Yes	24 months (Baseline, 6, 12 & 24 months)	Range of between 1 and 35 years	Not clear
Rivero-Arias, O. et al. (2006)	286	Multi-centre randomised controlled trial	Yes	12 months (Randomisation, 2, 6 & 12 months)	6 to <12 wk = 23% 3 to <6mo = 25.2% 6 to <12 mo = 16.5% ≥ 12 mo = 35.3%	Mean = 41.7
Ratcliffe, J. et al. (2006) & Thomas, K. J. et al. (2005)	241	Randomised Controlled Trial	Yes	24 months (Baseline, 3, 12 & 24 months)	4-52 weeks	18-65
Schweikert, B. et al. (2006)	382	Randomised controlled trial	Not reported	6 months	≥ 6 months	Mean = 46.7
Willich, S. et al. (2006)	3,451	Randomised Controlled Trial	Yes	3 months (Baseline & 3 months)	≥ 6 months	Mean age (Acupuncture) = 49.8; control = 51.4
Greiner, W. et al. (2006)	Simulated for a cohort of 100 patients	Decision Analytic Model	Yes	1 year of treatment	“chronic”	Not specified
Witt, C. et al (2006)	2,388	Nonrandomised Cohort	Not reported	6 months (Baseline, 3 & 6 months)	≥ 6 months	Mean = 52.9
Tarride, J. et al. (2006)	Simulated for a cohort of 2000 patients	Decision Analytic Model	Yes	12 Weeks	Not specified	Not specified
Neighbors, D. et al. (2001)	Decision analytical model with sample size not stated	Decision Analytic Model	Yes	1 year	“chronic”	Not specified
Spiegel, B. et al. (2003)	Decision analytical model with sample size not stated	Decision Analytic Model	Yes	Lifetime	“Chronic”	60 year old patients

Table 4.5 Continued...

Author (Year)	Sample Size	Study Design	Sensitivity Analysis?	Time-period (inc. frequency of data collection)	Average time since onset of pain	Age Range (Years)
Taylor & Taylor (2005)	Decision analytic model with sample size not stated	Decision analytic model	Yes	2 Years & life time	“chronic”	Not specified
Kemler, M. & Furnee, C. (2002)	54 (36 to SCS with PT, 18 to PT)	Randomised Controlled Trial	Yes	1 year (Baseline, 1 day prior to implantation, 1, 3, 6 7 12 months following)	“chronic”	Not reported
Thompson, M. et al (2005)	Decision analytical model with sample size not stated	Decision Analytic model	Yes	24 hours (2, 4 & 24)	Not specified	Not specified
Niemisto, L. et al. (2005)	204	Randomised Controlled Trial	Not reported	2 years (Before randomisation, 5, 12 & 24 months)	“chronic”	24 – 46
Lissovoy, G. et al. (1997)	Simulated for a cohort of 1000 patients	Decision analytic model	Yes	60 months	Not specified	Not specified
Haas, M. (2005)	837 with chronic low back pain	Nonrandomised trial	Yes	12 months (3 & 12 months)	“chronic”	Mean approximately 40.8
Thomas, K. (2005)	759	Randomized Controlled Trial	Yes	2 years (Baseline and at 2 years)	Not reported	≥ 45 years
Goossens, M. et al (1998)	148	Randomised Controlled Trial	Not reported	pre-treatment; immediately post-treatment; 6 & 12 months	9 yrs, 10 months	18 - 65
Kumar, K. et al. (2002)	67	Nonrandomised trial	Yes	60 months (Every 6 months)	Not Reported	Not reported
Goossens, M. et al. (1996)	131	Randomised clinical trial	Yes	60 weeks (2 weeks prior to treatment; upon starting treatment; upon completion of 6 weeks of treatment; 6 & 12 months after termination of treatment)	mean pain duration = 10 years	18-65 (mean age = 44)

4.4 Discussion

It can be seen that just over half of the economic evaluations summarised in Table 4.4 focus on back pain. Two have a wider focus on general chronic pain, although in these two studies there is still a focus on particular treatments/interventions. Other pain types listed in Table 4.4 are: Migraine, reflex sympathetic dystrophy, neck pain, knee pain, arthritis pain, fibromyalgia and neuropathic pain. Those treatments/interventions which appear to have received the most attention so far have been: Physiotherapy/ physical therapy, acupuncture, educational/ cognitive behavioural treatments, and opioids.

Eleven of the papers selected for review report the results of a cost-utility analysis, four report results from cost-effectiveness analysis, two papers discuss cost-utility *and* cost-effectiveness, and in three of the studies cost-minimisation was used. No papers were found in which cost-benefit analysis was used⁶. Fourteen of the studies involved the collection of primary data on costs and outcome; these were the thirteen randomised and nonrandomised trials and one study in which a decision analytic model was used (Tarride, Gordon et al. 2006). The remaining six studies in which decision analytic models were used involved the use of costs and outcomes identified from searches of the available literature.

Where cost-utility analysis has been used in studies in which primary data has been collected, utility values were obtained using EQ-5D in four cases and through converting SF-36 scores to SF-6D in three cases. Outcomes used in cost-effectiveness analyses so far have included: pain intensity on a Visual Analogue Scale, change in scores from SF-12, and Oswestry Disability Index. In the case of one cost-minimisation analysis, no significant difference in outcome was found between intervention and control group in terms of utility values for health states defined by an adapted version of the McMaster Utility Measurement Questionnaire.

⁶ At least one paper has been published since the search was conducted (Haefeli, Elfering et al. 2008) which focuses on spinal surgery for low back pain. The authors note: "we are not aware of any cost-benefit analysis using the contingent valuation approach with willingness to pay in the field of spinal surgery" (p575).

The study by Goossens *et al.* (1998) was difficult to categorise, although utility values were obtained through the use of rating scale and standard gamble techniques, results were not reported in terms of cost per QALY. The authors report that both programmes including operant treatment are equally effective in improving utility and more effective than no treatment. The conclusion is that “adding a cognitive/relaxation programme of about US\$1251 per patient to the operant programme did not result in any additional short-term or long-term effects” (Goossens, Van Molken *et al.* 1998), which would seem to leave us with a cost-minimisation analysis.

Almost half of the studies took a societal viewpoint in terms of measuring costs (i.e. direct and indirect costs were included in the evaluation). Only four studies were conducted in the UK, three of which focussed on low back pain.

The shortest period over which outcome was assessed was 24 hours (Thompson, Gawel *et al.* 2005); a 24 hour time horizon was chosen by the authors in this case as this is the median duration of a migraine attack. The other studies reviewed consider pain types where the pain will be more continuous and where treatments will either be delivered over a greater length of time, or, in the case of interventions such as surgery, be aimed at delivering a more prolonged benefit. Migraine can be considered to be unique in terms of symptoms and treatment; for other types of pain a longer time period would seem to be more appropriate.

At the other extreme, the longest period over which outcome was assessed was a lifetime; a further two studies used a 60 month (five year) time horizon. Such long time horizons are likely to mean that collecting data is extremely expensive and that response rates might be very low. For practical reasons, a time horizon of five years may only be feasible when using decision analytic models. Kumar *et al.* (2002) do consider a five year time horizon, although results were extrapolated for the five year period, apparently by amortizing the costs of therapy and maintenance from one year. With the exception of Kumar *et al.* (2002), the time horizon for studies involving the collection of primary data

varies between three and 24 months. It is clear that in some cases the time horizon involves a treatment period and then a separate follow-up period, whereas in other cases the data was only gathered over the duration of the clinical trial.

In only two of the 14 papers reporting studies involving the collection of primary data (Goossens, Rutten-van Molken et al. 1996; Kemler and Furnee 2002) was there some form of justification for the type of economic evaluation adopted. Goossens *et al.* chose to measure the effects of the interventions in terms of utilities because a utility is “a single comprehensive outcome measure that reflects the value or preference that respondents assign to a particular health state” (p1248). As no significant difference was found between the utilities of the intervention and control group the authors simply report the costs of each, i.e. as results of a cost-minimisation. Kemler *et al.* (2002) state that “In the societal perspective, patients’ health states have to be valued by the general public, because everyone is not only a potential patient but also a taxpayer. Therefore, each patient’s EQ-5D ratings were transformed into utility (i.e. a preference score that the general public would give for the health state as indicated by the patient).”

Similarly, in only two papers (Thomas, MacPherson et al. 2005; Schweikert, Jacobi et al. 2006) did the authors seek to justify their choice of outcome measure. Schweikert, Jacobi *et al.* explain that the psychometric properties of the EQ-5D had been successfully tested in a pilot study in a similar population of patients with chronic low back pain and, furthermore, had twice previously been recommended for these patients. Thomas, MacPherson *et al.* state that the SF-36, the Present Pain Intensity Scale of the McGill Pain Questionnaire and the Oswestry Pain Disability Index had all previously been piloted with this patient population. SF-36 pain dimension was chosen as the primary outcome measure, as the pilot data and other studies had shown that this measure was responsive to change in patients with low back pain and showed evidence of the validity and responsiveness of the measure.

A majority of those studies involving the collection of primary data used more than one outcome measure. No authors acknowledged or discussed any limitations of the outcome measures used. Studies will be discussed in turn below according to the main outcome measure used/reported.

4.4.1 EuroQol EQ-5D

Kemler and Furnee (2002) report that mean pain intensity, measured on a visual analogue scale, was reduced by 2.7 units in the intervention group, of patients with reflex sympathetic dystrophy, whereas pain intensity increased by 0.4 in the control group ($p < 0.001$). And yet, the mean health-related quality of life score improved for both groups, by 0.22 and 0.03 respectively ($p = 0.004$). So, even given the small sample size for this study ($n=54$), the two differences in pain intensity and health-related quality of life are statistically significant.

It is unclear whether this finding is the result of some problem with the way that patients completed the EQ-5D, if there is some anomaly with the EQ-5D tariff values, or if, instead, patients failed to understand the visual analogue scale. The control group did receive treatment (physical therapy alone) and so it is also plausible that improvements in health-related quality of life were experienced as a result of this treatment other than a reduction in pain intensity; if this is the case it is interesting that other aspects of quality of life for chronic pain sufferers can improve in response to treatment, apparently independently of the level of pain itself.

Rivero-Arias *et al.* (2005) report no significant differences in utility at any of the follow-up points between their intervention (spinal surgery) and control groups (intensive rehabilitation) with chronic low back pain. Adjusting for a slight difference in utility values at baseline meant that there was a mean QALY difference in favour of surgery of 0.068 (-0.02 to 0.156, $p = 0.13$), although this difference is still not statistically significant. Given such a tiny difference in health-related quality of life, and one which was not statistically significant, it is not surprising that a non-significant difference was also found between the groups in terms of return to employment.

In a separate study, Rivero-Arias *et al.* (2006) found that despite utilities improving both at the two month follow-up and 12 months, the mean difference between groups following two different interventions was not statistically significant. This result was supported by scores from the Oswestry Disability Index, although patients in the intervention group reported significantly enhanced *perceptions* of benefit. It is not made clear how patient perception was reported/ measured, but there could be a number of reasons for this. First, if a VAS was used then, once again, there is the possibility that respondents misunderstood how to use the scale, or used it inconsistently. Second, there may have been something resembling a placebo effect with respondents feeling better for receiving a regular intervention as opposed to a one off meeting simply to receive advice. Third, there is also the possibility that respondents were reporting their perception in a way that was expected to please the researcher, almost as an expression of gratitude. Finally, it may have been the case that patients' perceptions were accurate and correctly reported and that they experienced some benefit that was simply not detected by the five dimensions on the EQ-5D or by those on the Oswestry Disability Index.

Schweikert (2006) noticed that health-related quality of life scores for patients with chronic low back pain improved similarly in both the intervention and control groups during treatment. After six months, health-related quality of life was lower in both groups than at discharge, with the decline being less pronounced in the intervention group. It was found that in those cases where complete data was available the difference in health-related quality of life between groups was statistically significant, although upon transforming EQ-5D scores into QALYs and after accounting for baseline differences *no* significant difference could be detected (change in utility was equal to 0.008, 95% CI 0.028, -0.0113; $p = 0.396$). Despite showing no significant difference in terms of health-related quality of life, there were slightly fewer days off work due to spinal problems in the intervention group compared to the control group after six months and the number had fallen compared to that reported at baseline, whereas in the control group there was an apparent increase.

Although days off work due to illness can be used as an indication of the level of disruption to patients' daily activities, as well as their financial situation, it would be conventional and correct to include productivity losses as a cost (an indirect cost), rather than as an outcome; indeed Schweikert et al. do include indirect costs, and so by considering days off work as an outcome as well there would appear to be double counting. Schweikert et al. (2006, p2519), when summarising their results state that "six months after rehabilitation, patients in the intervention group were absent from work an average of 5.4 days less than patients receiving usual treatment. ... The cognitive behavioural treatment showed lower indirect costs...". The impression given is that a positive improvement in outcome is achieved by the intervention, which is associated with lower costs; in reality, the outcome and the reduction in indirect costs are part of the same phenomenon. A reduction in the number of days off will obviously reduce indirect costs.

Despite the problems associated with using days off work as an outcome measure, it is still a useful piece of information. If some patients in the intervention group were able to go back to work this would suggest that they should have reported an improvement in usual activities, and may reasonably be expected to have improved mobility. The fact that no significant improvement was detected for the control group vis-à-vis the intervention group therefore appears to raise concern about the EQ-5D. It is not clear, however, how concentrated the reduction in days off was across the full sample group, and we have no indication of how good or bad respondents' quality of life was on the days on which they did not return to work. It is still possible, therefore, that the EQ-5D is giving us the more accurate picture here.

To summarise, it can be seen that the difference in health-related quality of life, as measured by EQ-5D, between intervention and control groups has, in all cases reviewed here, been modest and in only one paper did the authors report a difference in utilities that was statistically significant. Although other outcomes were measured, in all of the studies EQ-5D was the only measure of health-related quality of life used, and we therefore cannot compare the results obtained through using EQ-5D with those obtained through the use of any

equivalent or competing measure. It can be said that in the study in which the Oswestry Disability Index was used, results from this measure of disability seemed to support those from EQ-5D. In three studies, however, the results from EQ-5D appeared to contradict with more basic indicators such as pain intensity, days off work due to illness, and patients' own perceptions of benefit.

4.4.2 Short Form-36 & Short Form-12 Health Survey

Ratcliffe *et al.* (2006) converted SF-36 responses to SF-6D scores to report health-related quality of life for a group of patients receiving acupuncture and a usual care group for low back pain at baseline, 3, 12 and 24 months. A sizeable improvement in quality of life was reported for both groups over the 24 months. Although there was a slightly higher mean health-related quality of life score for the acupuncture group at 24 months (74.2 compared to 72.9 for the usual care group), the mean differences between the groups were not statistically significant at any of the follow-up points. The difference in the QALY gain between the two groups was 0.029 (95% CI -0.034 to 0.092); again this was not statistically significant.

In the paper by Thomas, MacPherson *et al.* (2005) more detail can be found about this same study. Thomas, MacPherson *et al.* report results from use of the SF-36 pain subscale with the group of patients with low back pain. Only when those respondents who were permanently unable to work were excluded did the research team find a difference between the control and intervention group in terms of SF-36 pain scores. The estimated effect of the intervention at 12 months was 7.0 points on the scale. The effect was more sizeable at 24 months. Four other dimensions on SF-36 (Role-Physical, Physical Functioning, Social Functioning, and Role-Emotional) showed clinically significant change in both groups, smaller change was observed in the Vitality/Energy dimension. The between-groups effect for change between baseline and 12 months did not reach statistical significance for any of the SF-36 dimensions other than Pain.

Witt *et al.* (2006) did find that improvement in health-related quality of life, as measured by SF-36, was significantly more pronounced in a group of patients

receiving acupuncture for low back pain than in a control group. At three month follow-up, the difference in the physical component score on SF-36 was 4.7 ($p < 0.001$), and on the mental component score 2.1 ($p < 0.001$). In this study, the control group simply consisted of patients whose acupuncture was delayed. It is unclear what type of treatment, if any, the patients in the control group received during the period in which they waited for acupuncture; depending on the nature of this treatment, it is perhaps not surprising that a significant difference between the groups was found in terms of health-related quality of life.

SF-36 was also used by Willich *et al.* (2006), this time with patients with chronic neck pain. The authors report their results after three months as QALY scores; a very small difference (0.024) in mean QALYs is reported between the acupuncture and control group, although this difference is statistically significant ($p < 0.001$).

Haas *et al.* (2005) report improvement in health-related quality of life (measured by SF-12) alongside improvement in pain (measured on a visual analogue scale) and disability (Oswestry Disability Index). Improvement over time within all groups in terms of pain, disability, and health-related quality of life was statistically significant, and improvement in pain (and disability was significantly greater for the intervention group. The authors note, however, that there was little difference in the level of improvement between the intervention group and control group in terms of mental and physical health, measured with SF-12. For example, the score for both groups on mental health at 12 month follow-up was 4.9. There was a marginally better score for physical health in the control group at 12 month follow-up (14.7 compared to 15.8).

To summarise, in three of these four studies there is a reported improvement in health-related quality of life in response to treatment. In three studies using SF-36, one reported that no significant difference in health-related quality of life was found between intervention and control groups, one found a significant difference in QALY scores and one found a significant difference only in *some* components of SF-36. In the study by Thomas *et al.*, whereas no significant

difference was found in terms of the overall health-related quality of life score, there was a significant difference on the pain dimension on SF-36. Haas *et al.* also found significant differences between groups on some dimensions of SF-12 (including pain), but not on others.

4.4.3 Other outcome measures

Niemisto *et al.* (2005) used four outcome measures: pain intensity on a visual analogue scale, frequency of low back pain, Oswestry Disability Index, and the 15D⁷ quality of life instrument. Within the two year follow-up period, significant difference appeared between the intervention and control groups in pain intensity, but not in disability. There was a similar reduction in daily low back pain in both groups. The annual number of sick leave days decreased on average by five days during the two year follow-up period in the control group, while the decrease was two days in the intervention group. Health-related quality of life increased equally in both groups without any differences on any of the 15D dimensions between the groups. Health-related quality of life was *not* used in the cost-effectiveness analysis because of the high percentage of missing baseline data from 15D.

The paper by Niemisto *et al.* (2005) reports indirect costs and then further discusses annual number of sick days as an outcome, although incremental cost-effectiveness analysis is only reported for pain intensity and disability. Treating sick days as a measure of outcome has, as was discussed earlier, the potential to cause confusion and comparing sick days with costs would mean double counting.

Goossens *et al.* (1996) use the Maastricht Utility Measurement Questionnaire. This questionnaire requires patients to first rate their functional level over the past two weeks using a five-point scale for six domains: physical state and mobility, self-care, emotions, leisure activities, pain and other complaints, and side-effects of treatment. Patients then value their own health state, together

⁷ The 15D is a generic, 15-dimensional, self-administered instrument for measuring HRQoL among adults (Sintonen 2001).

with a reference state, using both rating scale and standard gamble. Participants in this study were divided into three groups: ECO, EDI and WLC (see Table 4.4). In general, there was no change over time and no difference between the groups on any of the six domains, except for significant differences during the treatment period in favour of the ECO (intervention) group compared to the WLC (waiting list) group on the leisure activities domain, and in favour of EDI compared to WLC on the self-care domain. Based on rating scale results, EDI treatment produced a QALY gain of 0.027 per patient compared to ECO treatment. SG utilities showed that EDI treatment yielded a utility gain of 0.022 QALY compared to ECO treatment. Due to the way the figures combine results from different patients at different moments statistical significance could not be tested.

The second study by Goossens *et al.* (1998) is very similar to the first. Patients were first asked to rate their current functional level using the five-point scale over six dimensions. The scores on the six dimensions of health did not differ between the groups. The actual valuation of their health state was then measured by means of rating scale and standard gamble techniques – both completed by the patients themselves. The economic evaluation confirmed that operant treatment for patients with chronic low back pain was more effective at improving utilities than no treatment. Adding a cognitive/relaxation programme did not result in any additional effects, but did result in an additional cost.

4.4.4 Studies in which parameters were identified through searches of the literature

In one study (Neighbors, Bell *et al.* 2001), assessing the cost-effectiveness of an opioid analgesic for the management of moderate to severe (general) pain, the authors use a value of 0.47 to represent the disutility associated with uncontrolled pain. The value was taken from a study which elicited preferences for a range of health outcomes associated with shingles-related pain in two American populations aged 65 to 70 years old (Bala, Wood *et al.* 1998). The same study by Neighbors *et al.* assumes the same disutility, or utility decrement of 0.1 for four separate side-effects from the medication: respiratory depression,

nausea/vomiting, mental cloudiness, and chronic constipation. Explanation as to the origin of many of the model parameters is given by the authors in a Table (Ibid. p135), a section of which is reproduced below in Figure 4.2. It can be seen that the source of many parameters is given as “assumption”.

Figure 4.2 Parameters Included in the Model by Neighbors *et al.*

Global Durations of Events, Disutility Values, and Unit Costs by Phase

Description	Value	Source
Disutility Values		
Titration	0.47	Assumed the same as uncontrolled pain
Withdrawal		
Stabilization		
Respiratory depression	0.1	Assumption
Nausea/vomiting	0.1	Assumption
Mental Cloudiness/Somnolence	0.1	Assumption
Chronic constipation	0.1	Assumption
Long-Term Use		
Noncompliance	0.47	Assumed the same as uncontrolled pain
Aberrant drug-related behavior	0.2	Assumption
Overmedication	0.2	Assumption
Need for increased dose	0.47	Assumed the same as uncontrolled pain
Other		
Uncontrolled pain	0.47	Bala et al developed a mean utility score for persons (n=114) with severe pain from shingles using the standard gamble method.
Controlled pain	0.27	Bala et al

Greiner *et al.* (2006) use utility weights for ‘controlled’ and ‘uncontrolled’ pain which are taken from a study involving patients with chronic low back pain. It is not specified by the authors how the values were elicited in the original paper and after tracing the original paper via the reference (Goossens, Vlaeyen et al. 1999), it is not clear how Greiner *et al.* have identified the values from it that they have used. Greiner *et al.* use a utility decrement for nausea and vomiting (side-effects from opioid treatment) that was elicited within a study of pregnant women (0.113). The authors assume that the utility decrement for mental cloudiness is the same as that for nausea and vomiting and derive a utility decrement for severe constipation (0.165), as no relevant values could be identified from the literature (Ibid. p294).

It is not clear how accurately the utility values used in some of the decision analytic models will reflect the utility of patients. Some studies appear to rest on three assumptions: (i) utility decrements for specific symptoms are transferrable between different contexts and different conditions; (ii) the utility loss arises

from the presence of the symptom and not from the effect it has on wider quality of life; (iii) utility decrements are independent and hence the decrement for each symptom present can be added together to find the total (which should be subtracted from one). Starting with the first of these assumptions, severe back pain may be undesirable because it restricts mobility and independence for long periods of time. Severe pain associated with arthritis may restrict mobility to a lesser degree but be tiring, with pain also persisting over longer periods. Shingles is an acute condition, although the pain itself, often described as being burning or gnawing pain, is likely to be deeply unpleasant until the condition is successfully treated. If the experience of pain itself is different in each of these cases, it may be that the utility decrement associated with each is different. Furthermore, it is not clear whether it is appropriate for utility values for 65 to 70 year old patients to be generalised to represent those of the general patient population with chronic pain. Older patients may, for example, have lower expectations of the quality of their 'normal' health and are more likely to experience co-morbidities.

In terms of the second assumption, 0.1 and 0.113 may be accurate reflections of the utility loss of nausea itself as a symptom/unpleasant experience. But, if a patient chooses not to drive, work or socialise when they have nausea then the presence of this symptom, as well as being unpleasant, could have quite a large impact on the quality of the patient's day. Hence the presence of the symptom would be associated with a larger utility decrement, the decrement associated with the unpleasant experience and that associated with loss of function, independence, loneliness, etc. The addition of the symptom mental cloudiness, while adding another unpleasant experience, is unlikely to impact as significantly because the patient is already unable to drive, work or socialise.

If the above example is observed in reality then it should be expected that the 'true' utility decrement (that reflecting the total accumulative impact of the symptoms) of many symptoms will be less than the arithmetic total of the utility decrements of each symptom individually. The opposite could equally be true. So, for example, the presence of one symptom may be easily tolerated, whereas nausea and mental cloudiness together are not, in which case the

result would be a larger utility decrement for both together than the sum of the individual decrements. This may indicate that an additive model is inappropriate, but it is also a problem caused by considering specific, micro-level indicators of health rather than broader themes relating to quality of life.

4.5 Summary

Twenty studies were found for review, 11 of which focussed on back pain. Fourteen involved the collection and/or use of primary data. The overall quality of the set of papers reviewed in this chapter could have been improved if all had given a clear description/ explanation of: the form of economic evaluation used; the types of cost included; the alternative interventions being compared; characteristics of the study participants. Likewise, overall quality would have been improved had the authors, where applicable, included a justification of: the type of economic evaluation used; the choice of primary outcome measure; the choice of time horizon.

In terms of the questions asked at the beginning of this chapter, these are addressed in turn below:

- 1. What types of economic evaluation have been conducted and what outcome measures have been used?*

The majority of studies involved the use of CUA. Four involved the use of CEA, in two both CUA and CEA were conducted. No examples of CBA were found. In only two papers was there some justification for the form of economic evaluation used.

Of the 14 studies in which primary data was collected and/or used, five used EQ-5D as the primary outcome measure, three used SF-36 with scores being converted to SF-6D, and three used pain intensity (or relief) as measured on a VAS. Several other questionnaires were also used.

2. *On what basis have researchers chosen and justified the use of a particular outcome measure?*

In only two of the twenty studies (21 papers) reviewed did the authors seek to justify their choice of outcome measure(s). Schweikert, Jacobi *et al.* explain that the psychometric properties of the EQ-5D had been successfully tested in a pilot study in a similar population of patients with chronic low back pain. Thomas, MacPherson *et al.* state that the SF-36 pain dimension was chosen as the primary outcome measure, as the pilot data and other studies had shown that this measure was responsive to change in patients with low back pain. In other words, the outcome measures were chosen because they were deemed to be good at measuring what they claim to measure. There appeared to be no consideration given as to the appropriateness of what was being included as an outcome, although there was a concern that the outcomes measured should be able to be somehow converted to a utility value.

While it may have become a convention (encouraged by organisations such as NICE) to use certain generic measures of health-related quality of life, and while the use of the 'off the peg' utility values associated with them is a simple option for researchers, this is not to say that they are 'fit for purpose'. There appears to have been little consideration of this important question.

3. *Does it appear as though some attempt has been made to detect broader changes in quality of life, or have studies simply assessed pain and disability? To what extent have those dimensions identified in chapter two been included?*

Most assess health related quality of life, with either the EQ-5D or SF-36. Of the likely affects of chronic pain listed in Table 2.2, and again in the introduction to this chapter, EQ-5D has dimensions relating to: Pain, activities of daily living, and psychological and emotional distress. Dimensions partially addressed by EQ-5D are: independence (self-care), employment (usual activities), and social issues (usual activities). The SF-36 covers more dimensions from this list

including: pain, activities of daily living, employment, social issues, and enjoyment of life. The SF-6D partially includes: independence (bathing and dressing) and sleep disturbance (tiredness).

4. *If more than one outcome measure has been used, is there a significant difference in the level of responsiveness and/or differences in the apparent desirability of an intervention as indicated by the different measures?*

All studies involving the collection of primary data reported some significant positive response to treatment. Significant differences were also found between groups receiving different treatments in terms of pain intensity and perception of benefit, although it was less common to find significant differences between groups in terms of health-related quality of life. In only one case in which the EQ-5D was the primary basis for utilities were significant differences found between the intervention and control groups. There is no way of knowing if it was the measures of health-related quality of life used which were not sensitive enough to detect differences between intervention and control groups, or whether it was simply the case that the treatments being compared were similarly effective.

It would certainly seem from the papers reviewed in this chapter that we can expect significantly different results when we use a health-related quality of life measure than when we use more basic indicators, such as pain intensity, or patient perception of benefit. It is unclear which approach is more reliable. Pain is also one single symptom; there is no reason to believe that a change in pain intensity just large enough to be statistically significant will lead to a significant change in more general health-related quality of life.

We see from the study by Thomas, MacPherson *et al.* (2005) that SF-36 was sensitive to changes in dimensions relating to participation, suggesting that a broader instrument can cover aspects of quality of life important to patients, while remaining sensitive to the type of changes expected from treatment.

5. *Over what timescales have results been collected? Do some instruments begin to detect changes in outcome with a shorter delay (lag) than others?*

The time horizon varied between 24 hours and the patient's lifetime, although both of these extremes are special cases. The timescale of 24 hours was for a study looking at migraine and the short timescale is unlikely to be appropriate in other contexts. The studies in which results were estimated over the patient's lifetime involved parameters being identified from the literature and included in decision analytic models. Other timescales used included six months (two studies), 12 months (six), and 24 months (five). Where results were compared at three months (Willich, Reinhold et al. 2006) on SF-36 only a very small difference was found between health-related quality of life in the control group and the intervention group. Thomas, MacPherson *et al.* (2005) measured health-related quality of life in a control and intervention group at intervals of three, 12 and 24 months. Although quality of life improved over the 24 month period, they found no statistically significant difference between the two groups at any of these points.

It was common, however, to find that intervention groups experienced statistically significant reductions in pain intensity than control groups, even where no significant differences in health-related quality of life could be detected between the groups.

4.6 Conclusion

It was noted in chapter three that there are a number of methodological issues with cost-benefit analysis and it was acknowledged that there are a number of reasons suggested in the literature as to why it CBA may be unsuitable for use in a health context. It is perhaps unsurprising then that no studies were found in which CBA was used.

In none of the papers identified was there an indication of which outcomes are important to consider in economic evaluations relating to chronic pain. Instruments were chosen which were valid and reliable, but the appropriateness

of the type and range of dimensions included in the instruments was not discussed.

All instruments designed to assess health-related quality of life were sensitive enough to detect improvements following treatment, but in only a small number of cases were sizeable and significant differences in HRQL detected between patient groups. It appears that from the instruments currently in use, SF-36 is more appropriate for use with chronic pain patients than EQ-5D. The SF-36 covers more of what is likely to be important to patients and appears to be more sensitive. There are still aspects of life which are likely to be important to patients which are not included in or addressed by SF-36.

It is acknowledged that further studies have been published since the review was completed in February 2008, and notable among these is the study by Haefeli, Elfering *et al.* (2008), in which a cost-benefit analysis was conducted to assess the economic effectiveness spinal surgery for low back pain. The study design was a CBA feasibility study using a CV survey with ex post willingness-to-pay/willingness-to-accept (WTP/WTA) questions. A total of 115 patients after lumbar fusion, discectomy, or decompression were asked to respond to an ex post questionnaire on their WTP/WTA for their respective intervention. Even if this study had been found in the search, however, it is not clear whether it would have fulfilled the inclusion criteria as it appears to include respondents with chronic and acute pain, and the results are not reported separately.

Mean stated WTP in the study by Haefeli, Elfering *et al.* was 20% lower than the actual operation costs (not known to respondents) for spinal fusion, although it was 37% higher for discectomy and 10% higher for decompression. The individuals' financial situation was the strongest predictor for WTP. Pain improvement, present pain, duration of hospitalization, and estimated intervention costs were significant independent predictors in the expected direction for the WTP, having controlled for sociodemographic and financial confounding variables.

In open-ended WTP questions the authors reported achieving response rates of around 71%. It is reported that reduced acceptability of WTP questions was partly explained by the personal nature of those questions, objections to the hypothetical nature of the questions, and by the fact that 9.6% would have paid “everything or as much as necessary” for surgery and therefore did not note a specific amount of money. The authors conclude that although CBA is shown to be feasible, refinement of the methodological approach would be necessary for any future study (Haefeli, Elfering et al. 2008, p585). It would appear – as was a concern expressed in chapter three – that budget constraints have played a significant role in determining results in the study by Haefeli, Elfering *et al.* The authors do not state whether any respondents expressed ‘protest votes’; this is different from a non-response.

CHAPTER 5: The Capability Approach

It has been suggested in earlier chapters that health is only one important aspect of quality of life and that a condition such as chronic pain can be expected to impact on almost every aspect of wider quality of life. It has also been suggested that because of the focus on health as the evaluative space, current forms of evaluation such as cost-utility analysis cannot be used to assess interventions falling outside of the direct scope of the healthcare system. One way of broadening the focus beyond health-related quality of life and of developing a new instrument for assessing outcomes associated with health or social interventions is to adopt an approach which is distinctly different from the traditional quality adjusted life year used in cost-utility analyses. The Capability Approach, originally developed and advocated by Amartya Sen, has now become recognised within development economics as a mainstream alternative to both income and utility based measures of well-being and quality of life. The approach is well known for providing the theoretical grounding for the United Nations' Human Development Index (United Nations Human Development Programme 2006). The strength of the approach, which focuses on the ability of individuals to lead a life that they have reason to value, is its wide evaluative space compared to rival approaches, which reduce well-being to a single dimension (health, utility or income).

This chapter begins by outlining the theory behind the Capability Approach (section 5.1). Focus then progresses to the methodological steps which will be involved in operationalising the approach, involving the choice as to whether to work with capabilities or functionings (section 5.2), the selection of capabilities in order to form the evaluative space (section 5.3), intrapersonal aggregation within the Capability framework (section 5.4), and finally weighting capabilities (section 5.5). Section 5.6 presents the recommendations to explore the Capability Approach within Health Economics and debate relating to similarity between capability and current practice in health economics. Section 5.7 summarises.

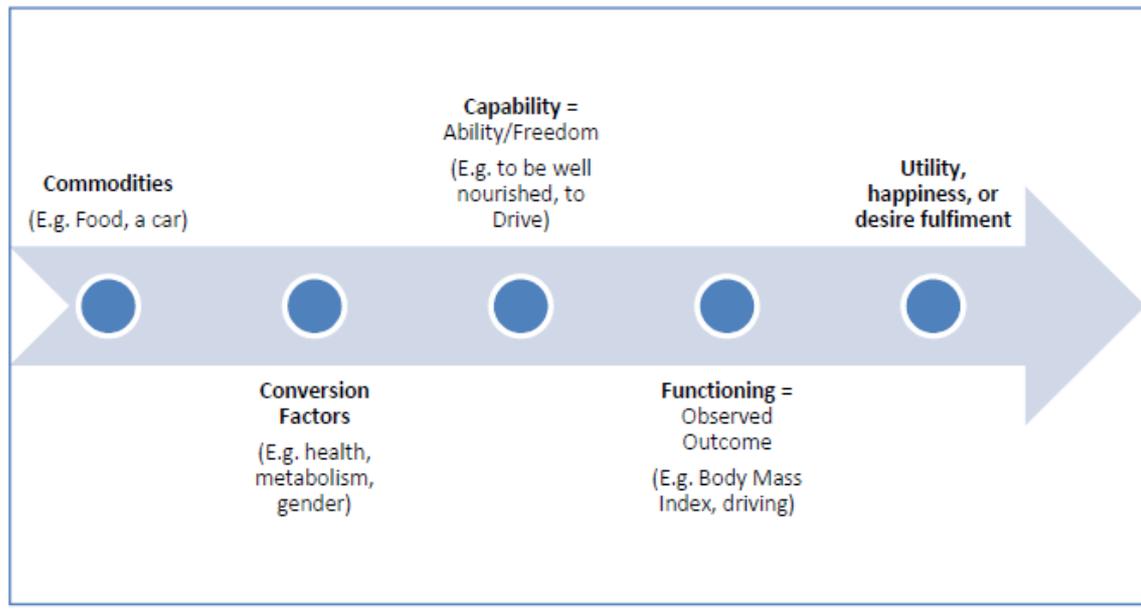
5.1 The Capability Framework

There is an important distinction within the capability framework between functionings and capabilities. Sen defines functionings as: the various things that a person manages to do or be in leading a life (Nussbaum and Sen 1993, p31). Some functionings are elementary, such as being in good health, working or being literate, while others may be more complex, such as achieving self-respect.

Capability is defined as the alternative combinations of functionings a person can achieve, and from which he or she can choose one collection (Nussbaum and Sen 1993, p31). See Figure 5.1 for an illustration of the relationship between commodities, capabilities, functionings and utility. This relationship will be described in more detail later in the chapter. Alkire (2005) defines capability as “a set of real opportunities that you could use in one way or another, the paths that lie open before you” (p121). By ‘real opportunity’ we mean the freedom or opportunity to accomplish what we value; Sen is very clear that this does not include freedoms or opportunities that a person might hold theoretically or legally but that, in reality, lie well beyond their reach (Alkire 2005, p121).

Section 5.1.1 outlines how the evaluative space is defined and made up under the Capability Framework. Section 5.1.2 compares this evaluative space to that used in utility-based approaches and section 5.1.3 to John Rawl’s idea of primary goods.

Figure 5.1: Commodities, Capabilities and Functionings



5.1.1 Specifying the Evaluative Space

In order to use the Capability Approach in an evaluative exercise one is initially required to specify the 'evaluative space'; establishing the objects of value? In standard utilitarian analysis the evaluative space would consist of the individual utilities. The capability approach sees the evaluative space in terms of functionings and capabilities to function. Sen regards the selection of the evaluative space to have "a good deal of cutting power on its own, both because of what it *includes* as potentially valuable and because of what it *excludes*" (Nussbaum and Sen 1993, p33).

There is also a second question: how *valuable* are the respective objects? Sen notes that answering the initial question on the identification of the objects of value does not, on its own, yield an answer to the second question regarding their relative values (Nussbaum and Sen 1993, p33); the requires a further evaluative exercise. It is further noted that individuals may differ a great deal from each other in the weights they attach to different functionings, however valuable they may all be.

Sen calls for two further distinctions, which, in turn, leads to a fourfold classification of points of evaluative interest. One distinction is between (1.1) the promotion of the person's *well-being*, and (1.2) the pursuit of the person's overall *agency goals*. The second distinction is between (2.1) *achievement*, and (2.2) the *freedom to achieve* (Nussbaum and Sen 1993, p35). This establishes four different concepts of advantage, which are set out in Figure 5.2.

Figure 5.2: Value Purposes

	Well-being	Agency Goals
Achievement	Well-being Achievement	Agency Achievement
Freedom to Achieve	Well-being Freedom	Agency Freedom

The terms in Figure 5.2 require careful explanation. Well-being is a narrower concept here than agency success; the important distinction is that agency goals are not restricted to the advancement of the person's own well-being. Well-being refers simply to the person's own state. A person's agency goals could be altruistic in that the individual may value functionings which benefit family, friends or society in general. Agency freedom is freedom to achieve whatever the person, as a responsible agent, decides he or she should achieve (Alkire 2005b, p219).

It is suggested by Sen that in determining whether a person is deprived in a way that calls for assistance from the state (e.g. overcoming hunger or illness) a person's well-being may arguably be more relevant than his agency success (Nussbaum and Sen 1993, p36). Furthermore, "for adult citizens, well-being freedom may be more relevant to state policy, in this context, than well-being achievement" (Ibid, p36). The state has a duty, in this case, to provide the individual with adequate opportunity to overcome their hunger and it is this opportunity, or freedom, that is important rather than the actual outcome. Robeyns (2006) discusses this issue with reference to a particular principle which has been labelled "luck-egalitarianism". Under this principle each person should have the same real opportunity (capability), but individuals should also be held responsible for their own choices. Clearly, if one wants to implement

this principle then the appropriate specification would involve a focus on capabilities, rather than functionings.

A further example may be useful in stressing the difference between well-being and agency. Alkire (2005) uses the example of a riverside picnic being interrupted by the chance to rescue someone from drowning. In this case your agency freedom (and hopefully achievement) increases, because you can save a person's life; but your achieved well-being diminishes, as you emerge cold, wet and hungry. It is not difficult for us to consider how this concept of well-being and agency might apply to individuals with chronic pain. An important theme that was found to run through chapter two was relationships with friends, family, carers and the medical profession. We could reasonably say that an individual with chronic pain who chooses not to ask friends or family for help because of a reluctance to be a 'burden' is exercising their agency freedom; but by not receiving help their own well-being achievement is diminished. Alternatively, the individual may feel they have no choice but to seek help; in this case their agency freedom is diminished, but through receiving help and support their well-being achievement increases.

It has been suggested that the state should focus on well-being freedom and direct policy so that individuals are provided with 'adequate opportunity' to overcome difficulties. It has also been explained that this would make capability, rather than functionings, the relevant focus. A further reason for considering capability is that freedom may have intrinsic importance for the person's well-being achievement (Nussbaum and Sen 1993, p39). The argument follows that, if choosing is seen as a part of living then, "even 'well-being achievement' need not be independent of the freedom reflected in the capability set. In that case, both 'well-being achievement' and 'well-being freedom' will have to be assessed in terms of capability sets" (Ibid. p39).

5.1.2 Happiness & Desire Fulfilment

It is the wider informational base that sets capability aside from standard utility-based approaches. There is no insistence within the capability framework that

only happiness be valued. Rather, the state of being happy is one among several objects of value. Likewise, there is no insistence that focus should only be on desire fulfilment; desire is regarded by Sen as “useful but imperfect evidence – frequently distorted – of what the person herself values” (Nussbaum and Sen 1993, p48). Chapter three contained discussion relating to utility as both desire fulfilment and as happiness. A central feature of Sen’s argument against the use of utility – interpreted in this way - is that valuing is not the same thing as desiring, and that the strength of desire is influenced by considerations of realism in one’s circumstances (Sen 1999, p14).

5.1.3 Primary Goods

While developing his Capability framework as an alternative to standard utility-based approaches Sen has also fought an argument on a second front. When used to make comparisons (i.e. when used as a means of viewing personal advantage and/or social deprivation) the capability framework finds itself in direct competition with commodity-based approaches, and in particular Rawls’ primary goods approach. Under the heading of ‘primary goods’ Rawls includes not only commodities but other goods such as liberty and self-respect. Sen states clearly that a focus on capability differs “*both* from the utilitarian concern with just mental reactions and from Rawlsian concern with primary goods” (Sen 1983, p163); this distinction is illustrated in Figure 5.1. It is argued that the Capability Approach continues to focus on human beings, rather than “moving with Rawls to incomes, goods and characteristics” (Ibid. p164). The problem, as Sen sees it, with primary goods is that the commodity requirements of a capability fulfilment not only vary between one community and another but that there are also differences within a given community in the ‘mapping’ of commodities to capabilities (Ibid. p164).

Robeyns (2005) argues that the relation between a commodity and the functionings to achieve certain beings and doings is influenced by three groups of ‘conversion factors’. *Personal* conversion factors (e.g. metabolism, intelligence, or gender) influence how a person can convert the characteristic of the commodity into a functioning. *Social* conversion factors (e.g. public policies,

social norms) and *environmental* conversion factors (e.g. climate, geographical location) also play a role in the conversion from characteristics of the good to the individual functioning (Robeyns 2005, p99). Hence, knowing what goods a person owns or has access to is not sufficient to know which functionings he can achieve. Consider, for example, a car, purely as a means of transport. If, due to the side effects of medication for severe pain, an individual is unable to drive then without help from another driver the car is utterly useless in enabling the functioning of mobility for this particular individual.

Certainly Sen and Robeyns would not claim that the availability of commodities, social institutions, economic growth and technical performance are not important, but the Capability Approach presses the point that they are not the ultimate ends of well-being. In this respect, there is some similarity with the literature on experienced utility introduced in chapter three, although there are other fundamental differences between the two approaches.

5.2 Operationalising the Capability Approach

So far, the capability approach has been explained and attempts have been made to justify its use above alternative approaches. One thing that has not been explained is exactly how, and indeed if, the capability approach can be operationalised. It should be realized that the theory is often described as 'incomplete'. Robeyns notes that the capability approach in practice comes in a variety of forms, in part because of the wide scope of the approach, but also because the approach is "radically underspecified" (Robeyns 2006, p353).

Sen has recently acknowledged that "there are widespread doubts about the possibility of making actual empirical use of this richer but more complex procedure" (cited in Robeyns 2006, p353). Two quotes can be offered to give a flavour of these 'widespread doubts'. Rawls has simply labelled the capability approach "an unworkable idea" (Rawls 1999, p13), and Sugden has argued that:

Given the rich array of functionings that Sen takes to be relevant, given the extent of disagreement among reasonable people about the nature of the good

life, and given the unresolved problem of how to value sets, it is natural to ask how far Sen's framework is operational. (Sugden 1993, p1953)

As a starting point for operationalising the approach Robeyns (2006), having searched the literature, has identified three important theoretical specifications: (i) the choice between functionings and capabilities, (ii) the selection of relevant capabilities, and (iii) the issue of weighting the different capabilities for an overall assessment (indexing). Each of these three steps will be considered in more detail in the following three sections of this chapter.

5.2.1 Functionings or Capabilities?

The choice between capabilities and functionings has been discussed already in this chapter and it has been argued that it may not be sufficient to restrict oneself to a consideration of functionings alone. On the other hand, even when we are ultimately concerned with capabilities, informational problems or measurement constraints may mean that a focus on functionings is the only feasible option. In 'The Standard of Living' (Sen 1987), Sen recognised that the data requirements for measuring capabilities would be very great. In *Inequality Re-examined* (Sen 1992) the possibility that capabilities might be measured if sufficient data were available is dropped. An evaluation of well-being achievement, we are told, should consider a person's capability set, but "this cannot be directly observed and must be surmised from empirical measures of that person's achieved functionings" (Alkire 2002, p181). It should be noted that Sen's comments on this issue relate to well-being achievement, the most basic, or limited, space in which we can assess personal advantage.

One reason for this pessimism might be that in the context of overseas development the macro-level data which is available to use is often both limited and has not been collected for the purpose of assessing capability. For example, the dimensions contained within the Physical Quality of Life Index (PQLI) were selected largely on the basis of the availability of data (Morris 1978). Data collected within developing countries may also be incomplete and the cost of collecting additional data significant. In a health context, at least

within the UK, data is routinely collected directly from the patient during studies aimed at assessing the cost-effectiveness of health care treatments. Most current health instruments can be thought of as assessing functionings. If it is possible, there may be more scope in the context of health – because of the relative ease of data collection – to attempt to assess capability than there is in the context of overseas development.

Where then does this leave one who wishes to move beyond well-being achievement? How are we to assess either well-being freedom or agency freedom? If it is capability that we seek to observe then it may be possible on a small-scale simply to ask individuals or groups whether they are or would, theoretically, be able to achieve a range of capabilities. If this method was used, however, one may question how reliable the response(s) would be! As an example, an individual in a wheelchair may physically be able to get to and enter a particular shop because the law states that the individual should be allowed to park near to the entrance and that there should be a ramp leading into the premises. The individual may also have a relatively high income. If, however, the individual has spent all of their disposable income for the month in which they are interviewed or, if their car is being repaired on that particular day, or for any number of possible reasons, the individual may well reply that they are unable to shop at this particular store. We may not regard this to be a useful answer, but at the same time we cannot claim that the answer is inaccurate. In short, we can never hope to know all of the considerations and 'background' behind such a response and we can never, therefore, be entirely sure that we are interpreting such a response correctly. If it is possible to directly assess capability then the difficulty is likely to be asking the right questions and asking them in an effective way (one which avoids confusion and which prompts an answer which can be interpreted accurately). This issue will be discussed further in chapter eight with specific reference to the empirical work conducted as part of this thesis.

Even if it is found that directly assessing capability is too troublesome and inaccurate then the approach may still be useful. It is worth reflecting on Sen's sentiments that functionings can be measured as a proxy for capability. The

first step would be to identify through qualitative work what people's agency goals are (providing for one's family, being loved, and so on). From this the researcher would have an idea of which functionings relate to these capabilities. Then observing that an individual achieves all of the related functionings would seem a good indication that they have high achievement in terms of the capability. Achievement of the functionings may be more objective than the respondent's subjective reporting of their achievement in terms of capability (how loved they feel for example). When using instruments to collect data from patients, achievement in terms of functionings may be less abstract/more practical for the patient to consider than the concept of capability.

Of course, there would be a need to identify the list of important capabilities in order to develop a list of functionings which are indicative of achievement in terms of these capabilities. In other words there is a need to know the overall goal before progress towards this goal can be assessed. When qualitative work was conducted with older people (Grewal, Lewis, et al. 2006) the researchers noted that respondents appeared to talk about their quality of life in terms of their ability to do and be things that were important to them. As noted above, many instruments used within health can be thought of as assessing functioning, but the process through which the functionings were chosen for inclusion is often not transparent. It is often only the perception/assumption of the research team that what instruments are measuring is what is important to patients. Indicators are used in the expectation that these provide some brief overview of a person's health related quality of life without first asking the questions 'what is quality of life' and 'what do patients regard as the most important aspects of it'.

5.2.2 The Selection of Relevant Capabilities

The second specification is the selection of relevant capabilities. On this issue there is a vast literature, the majority of which would prove a significant distraction from the intended direction of this thesis. Without wishing to become distracted by the intense debate surrounding the issue of lists, it is only right that there is a brief acknowledgement of both arguments and a justification for

favouring one position above the other. On the one hand, Nussbaum (2003) advocates the use of a fixed and central list of capabilities relevant to the evaluation of quality of life. In contrast, Sen has refused to endorse such a list. Nussbaum has a particular motive for developing such a list which relates to the specific task for which she has chosen to use the approach. Nussbaum seeks to use the capability approach to criticise moral norms. If she did not explicitly list the various capabilities that relate to some vision of the good life then she would be unable to criticise norms or practices which restrict these capabilities (Qizilbash 2002).

Sen has refused to give a definitive list of capabilities, largely because of the great variety of contexts in which the approach can be developed and used. With respect to the issue of lists Sen comments that: “What we focus on cannot be independent of what we are doing and why” (Sen 2004). He remains opposed to the “fixing of a cemented list of capabilities, which is absolutely complete... and totally fixed...” (Sen 2004); and insists that “pure theory cannot freeze a list of capabilities for all societies for all time to come, irrespective of what citizens come to understand and value”.

As Nussbaum’s list was designed to fulfil a specific purpose, distinct from assessing quality of life, it is not obvious that either her full list, or a sub-set of capabilities from it, should be used in this study to assess quality of life in patients with chronic pain⁸. Instead, within this study a new and context specific list of capabilities will be identified and used, something that Sen has never explained how to do beyond stating in general terms that some democratic process and public reasoning should be involved. This is not to say, however, that the process of selecting relevant capabilities has not been considered and we can look at this point to the work of both Robeyns (Robeyns 2003; 2005b) and Alkire (Alkire 2002; 2005).

⁸ Nussbaum’s List has been used by Lorgelly, Lorimer, et al (2008) and has formed the basis of a questionnaire for use in the evaluation of public health and social interventions. This work is ongoing.

Few studies so far have been conducted with the intention of identifying lists of valuable capabilities. If, when a greater number of studies have been conducted across a greater variety of contexts, it is found that the capabilities identified are strikingly similar then this will fuel the argument for relying on either one or a small number of generic lists. If it emerges that the type of capabilities reported as being valuable for a good life vary significantly according to the context in which the list is formulated, then selection of capabilities will need to be undertaken on the ground over and over again. At least for now, the *process* used for selecting capabilities is as important as which capabilities are selected and any process that is used should safeguard against selection bias. Robeyns suggests that the most likely source of selection bias will be the “life, values and social embeddedness of the researcher” (Robeyns 2005b, p203).

Robeyns (2003, p70) suggests that when selecting capabilities the following five criteria should be met:

1. *The criterion of explicit formulation:* The list should be explicit, discussed and defended.
2. *The criterion of methodological justification:* There is a need to clarify and scrutinise the method that has generated the list and justify this as appropriate for the issue at hand.
3. *The criterion of sensitivity to context:* The level of abstraction at which the list is pitched should be appropriate for fulfilling the objectives for which we are seeking to use the capability approach. For example, the list of central capabilities drawn up by Nussbaum may be relevant to discussion for national constitutions and Bills of Rights; its relevance for small-scale activities such as ours is constrained by the prescriptive character of the list as well as uncertainty regarding its ability to take account of individual and cultural differences.
4. *The criterion of different levels of generality:* If the specification aims at an empirical application, or wants to lead to implementable policy proposals, then the list should be drawn up in at least two stages. The first level will be an ‘ideal’ list; further levels would become more pragmatic. Distinguishing between the ideal and the ‘second-best’ level

is important because the second-best constraints might change over time, for example as knowledge expands or political systems change.

5. *Exhaustion and non-reduction*: The capabilities on the list should include all elements that are important. Moreover, the elements included should not be reducible to other elements.

Robeyns recognises that these are only general criteria and states that their only goal is to provide a sort of ‘check and balance’. Clearly a great deal more detail would be needed, particularly relating to the second criterion, before any researcher would be able to use Robeyns’ criteria as the foundation of a study proposal. In a later article Robeyns (2005b) goes some way towards providing an actual procedure for generating lists of capabilities. First she distinguishes between three categories of quality of life measurement, for which different procedures will be needed: small-scale projects, large-scale empirical assessments, and large-scale policy design. Because in this thesis the commitment is to assessment of quality of life within a relatively small group, discussion shall be restricted to the first category.

Small-scale projects are characterised by the fact that it is relatively clear who the affected persons are, and all affected persons can, in principle, meet to discuss the project (Robeyns 2005b). In such a setting Robeyns suggests that capabilities are selected through participatory methods and debated in a group. This position is shared by Alkire (2002), who believes that when selecting capabilities for the assessment of a small-scale project, one should support the principle that “the most local agent(s) capable of making a decision should make it” (p44).

The fact that a recommendation to select capabilities through ‘participatory methods’ is the most prescriptive guidance that can be found on this integral step towards operationalising the approach is further proof of how underspecified the approach is.

Furthermore, it should be recognised that a participatory approach is not without potential limitations. Nussbaum has argued that, just as people might adapt

their preferences or desires to deprived circumstances, people might adapt their list of capabilities to deprivation. Evidence of this can be found in chapter six where empirical studies in which the Capability Approach has been operationalised are reviewed.

5.3 Intrapersonal Aggregation

The third theoretical specification that is required for the application of the capability approach is the issue of aggregation, or indexing. The focus here will be intrapersonal aggregation, i.e. valuing an individual's capability for comparison with that of others. Two questions can be asked: (i) should capabilities be valued in such a way as captures tradeoffs between them?; and, if so (ii) how should we go about obtaining relative weights, or rather how should we capture this trade-off? Clearly, we may decide that it is unnecessary to weight at all; alternatively, if we object to assigning *relative* weights, then by default we must assign equal weights to capabilities or functionings.

5.3.1 Is Weighting Necessary?

Not all applications of the Capability Approach require intrapersonal aggregation, and in fact Robeyns warns that indexing may amount to a loss of information (Robeyns 2006). For example, if we wish to conduct a detailed analysis of the distribution of well-being within a particular population group then it will mean more to consider the individual aspects of well-being than blurring everything together into a single number. The question to be addressed here is: do we *need* to value the quality of life states defined by a capability questionnaire and/or quantify trade-offs between the capabilities? Within the context of health, if the objective was to formulate some form of 'capability-adjusted life year', it would seem that the answer to this question is yes. An alternative would be to leave the 'raw' data from the questionnaires as it is and restricting all economic evaluation to the approach of cost-consequence (Coast 2004; Coast, Smith et al. 2007a). Without some formal way of summarising the findings though it would be a near impossible task to consider and weigh up all of the aspects of quality of life which are affected by a project and detected by a

questionnaire against the monetary cost of the project. It would seem then that pragmatically some method of scoring responses to the questionnaire is required.

5.3.2 Equal Weights?

An argument in favour of treating every individual capability in a set as being equally important would be to say that a person is deprived if they fall down on any of the capabilities and cannot be compensated for a loss of capability in one area by having greater capability in another. Nussbaum (2003), for example, argues that “a society that does not guarantee [the capabilities on Nussbaum’s list] to all its citizens, at some appropriate threshold, falls short of being a fully just society”, she goes on to add that “although in practical terms priorities may have to be set temporarily, the capabilities are understood as both mutually supportive and all of central relevance to social justice. Thus a society that neglects one of them to promote the others has short-changed its citizens...” (p40).

Nussbaum’s argument is compelling when it comes to the most basic capabilities, such as having access to clean drinking water, food and shelter. It would be unlikely that making an even more plentiful supply of clean water available to a person could compensate them for sleeping on the street. When it comes to a broader list of capabilities, however, Nussbaum’s argument may prove to be somewhat idealistic. It could be argued that an individual with chronic pain makes choices and faces trade-offs between capabilities most days. In choosing to be free of pain an individual may effectively be choosing to take medication which has unpleasant side-effects (such as severe drowsiness) which, in turn, may mean that the individual is not able to drive or perhaps even leave the house. In choosing to walk about and to do shopping independently the individual may be accepting that later in the day they will experience pain and/or other forms of discomfort. While, ideally, society will find ways to enhance the capability of the individual across the entire capability set, there may often simply be no practical means of doing so. It could perhaps be said

that in reality we face Nussbaum's 'temporary' time period for the foreseeable future.

Recall, however, that Nussbaum's objective in using the Capability Approach is the criticism of certain moral norms (Qizilbash 2002) and not the evaluation of specific projects/treatments/social interventions. If one views a list of capabilities such as Nussbaum's almost as core human rights or entitlements, then it is more justified to claim that there should be no scope for tradeoffs; however our reasons for using the Capability Approach are distinctly different. In the case of evaluating health and social care interventions it appears that tradeoffs can be a very real and unavoidable part of everyday life, and that any process of evaluating such interventions should recognise and reflect the fact that some outcomes may carry greater weight/ benefit than others.

With the possible exception of basic human entitlements, it seems inevitable that individuals in a society will place more value on one particular capability than they place on another.

On a purely practical level, the problem with accepting the argument against trade-offs is that it makes interpersonal comparisons potentially impossible. Consider, for example, that we have identified three capabilities (A, B, and C) and their corresponding functionings (a, b, and c) as being of value in a particular context and wish to use them to assess the well-being of three individuals. We observe that Tom achieves the functionings a and b, while Dick achieves b and c, and Harry achieves only a. We can say that Harry has the lowest standard of living⁹, but we cannot make a judgement as to whether Tom is more or less deprived than Dick. Consider now that all three individuals regard A as being less valuable than C, then surely it would be wrong to insist that Tom is as well off as Dick. In this scenario it may be beneficial to allocate weights to the different functionings/capabilities, and so incorporate them into a single index.

⁹ Assuming that A is not more valuable than both B and C, but this would not be known if preference weights were not elicited for the three functionings or capabilities.

Of course, the example of Tom, Dick and Harry goes little way to demonstrating the enormity of the task of weighting capabilities, it was simply assumed that an individual either does or does not possess the capability to achieve a functioning. Whether or not Tom can achieve the functioning *c* may depend on how strict we are in our judgement. The point here is that we will always be setting an arbitrary standard in order to judge achievement, a sort of arbitrary 'pass mark'; but we do not need to restrict ourselves to using just 'pass' and 'fail', it may be more informative to consider the location of individuals along the full spectrum of achievement.

5.4 The Practical and Conceptual Complexity of the Weighting Task

There are two general approaches to weighting which have been used in the capability literature. The first approach simply involves allocating certain weights and then justifying them. An example of where this approach to weighting has been used is the Human Development Index (Qizilbash 2002; Chowdhury and Squire 2006). Some economists find this an arbitrary process and disapprove of the explicit value judgements involved (Robeyns 2006). The second option is simply to derive the weights in a statistical way, if the functionings themselves are measured using a statistical method. A third possibility could be to let the relevant group of people decide upon the weights themselves. Just like the *selection* of relevant capabilities, their *weighting* could be done through participatory methods; some form of social choice procedure.

This appears to lead us down the route of valuing functionings/capabilities according to individual preference, but can the weighting of capabilities based on preference be theoretically justified?

5.4.1 Happiness & Desire Fulfilment

Sen has been critical of using utility, defined as happiness or desire fulfilment, as a measure of well-being (see chapter three). The case that we are frequently asked to consider is that of two individuals, Person A is happier (or has more of his desires fulfilled) than Person B, but Person B is better off in

terms of functionings (e.g. being healthy and well-nourished). Sen would conjecture that considerations of feasibility and of practical possibility are entering into what Person A is daring to desire and what he is pained not to get; but this mental attitude does not wipe out the fact that this person is physically worse-off than Person B, and nor does it mean that he would not value an improvement in his condition. Thus Sen states that “valuation is a reflective activity in a way that ‘being happy’ or ‘desiring’ need not be” (Sen 1999, p19).

What Sen is objecting to is the idea of equating a person’s happiness (utility) with their overall quality of life. It may obviously be the case that the quality of a person’s life would determine their level of happiness, but what is also apparent within the capability framework is that we could still judge an unhappy person as otherwise having a “reasonable” or perhaps even “good” quality of life. We can fit happiness into our capability set, but to assess quality of life by relying on happiness (utility) alone would mean adopting an unduly narrow focus.

It has been suggested by Sen that considerations of feasibility and practical possibility influence what individuals dare to desire, and this suggestion appears to be based on a theory that individuals almost subconsciously seek to protect themselves from disappointment and therefore tend to limit their desires to what they feel they are reasonably able to obtain/achieve. This is not to say that they would not value achieving more than they had dared to hope for. Related to the discussion on desire, but subtly different is the issue of adaptation. Adaptation to one’s current circumstances may be associated predominantly with happiness. But adaptation is perhaps also related to desire. While disciplined and realistic desires (and therefore the avoidance of disappointment) will help to keep a person contented, his contentment must, by definition, be with his *current* and *actual* circumstances, which in turn are determining what is deemed to be realistic and hence what is desired. Happiness and desire, while different, seem to be inextricably linked via adaptation as it will only be after one has learnt to adapt one’s desires that one can be happy in one’s current state.

As already suggested in chapter three, it may be inappropriate simply to take the patient’s *utility* as an indication of their quality of life; the aim instead is to

value *quality of life* itself, described in terms of the capabilities covered in the questionnaire. It may also be appropriate to attach to capability states societal rather than individual preference values. As it happens, there is already precedence in health economics of obtaining values from the wider population, reflecting the reality that the general population pay for health services from their taxes and are all potential service users (Drummond and McGuire 2001). Furthermore, there is a common view advocated by philosophers such as Rawls, that decisions are most likely to be fair if they are made by people who do not stand to gain or lose from them (Gold, Siegel et al. 1996).

5.4.2 Utility as Choice

The third and altogether different interpretation of utility discussed by Sen is choice. Here, utility is the “real-valued (i.e. numerical) representation of choice” (Sen 1999, p12). If a person’s choice function has certain characteristics of internal consistency, then it can be represented by one binary relation and all the choices can be seen as maximizing according to that binary relation. The binary relation is frequently seen as utility.

Sen argues that: “whether the binary relation of choice can possibly be seen as reflecting the person’s well-being must rely on the motivations that underlie choice” (Sen 1999, p13). “Choice is obviously a very different type of activity from valuation, and in so far as it has a connection with valuation, this must partly arise from choice being a reflection of desire. Thus, much of what was said about the desire interpretation of utility will apply here too...” (Sen 1987, p12).

5.4.3 Scope for a little pragmatism

Having set out this minefield, it has to be acknowledged that Sen actually takes a surprisingly pragmatic view when it comes to valuation, as can be seen in the quote below, taken from *Commodities and Capabilities* (Sen 1999, p32):

Sometimes it might make sense to use utility-type information about strength of desire as reflecting valuation, even though the two are neither

identical, nor invariably closely related to each other. Sometimes even market-choice information can be used to arrive at weights to be attached to commodities and –from them- to guess the valuation of the corresponding functionings.

In all these exercises clarity of theory has to be combined with the practical need to make do with whatever information we can feasibly obtain for our actual empirical analysis.

It would seem then that in selecting a method for eliciting values for the capabilities we have relatively free rein, although in *Development as Freedom* (1999a) Sen does give further consideration to the issue of weighting capabilities, he writes:

... in arriving at an “agreed” range for social evaluation, there has to be some kind of reasoned “consensus” on weights, or at least on a range of weights. This is a “social choice” exercise, and it requires public discussion and a democratic understanding and acceptance.

It would seem that what Sen is saying is that utility (whether interpreted as happiness, desire or choice) should not be regarded as the object to be valued, but that in certain circumstances it may be the best indication that we have about the value of an alternative object; specifically in this case a functioning or capability. In this case we could ask respondents a number of different questions, such as: how happy would you be if you had this particular level of functioning? How much would you desire a certain degree of improvement in your level of functioning? Which level of functioning would you choose to have? In this sense, the valuation exercise would not be dissimilar to current methods familiar to Health Economists. What we have to decide is whether we wish to ‘settle for’ such methods/techniques as being acceptable, or explore further techniques which may be more theoretically appropriate. This decision will be made in chapter seven.

5.5 Capability and Health

After finding that standardised quality of life measures, such as the Short-Form 36 and the Nottingham Health Profile, did not cover all effects that were anticipated from a project looking to integrate care to the elderly, Verkerk *et al.* (2001) went in search of a more general framework of quality of life. The framework that Verkerk *et al.* recommend for future research is Amartya Sen's capability approach. It is recognized by the authors that Sen is interested in developing a concept of quality of life in general, not just health-related quality of life. It is stated by Verkerk *et al.* that: "therefore, his approach needs to be adapted if one would like to apply his ideas in this field of science" (Verkerk, Busschbach *et al.* 2001, p52). In order to operationalise the approach in the field of health-related quality of life the authors also see the need for compromise between what is acceptable on the grounds of practical difficulties of data availability, and what would be the right procedure had one not been limited in terms of information.

It is probably more accurate to say that the Capability Approach has the necessary *flexibility* to be applied to the field of health-related quality of life, rather than saying that the approach needs to be adapted. Flexibility is, as we have seen, already an integral and distinguishing aspect of the approach. Verkerk *et al.* do demonstrate an appreciation of the fact that the method of operationalising the approach will depend on the exact nature of the research question. The authors state that: "qualitative research is as necessary as quantitative research in order to determine which functionings should be included" (Verkerk, Busschbach *et al.* 2001, p52).

To 'adapt' the approach in order to apply it to the field of health-related quality of life might suggest that the list of valuable functionings is restricted to those relating directly to health, but this does not appear to be what the authors are advocating. Perhaps the concern is that to use the capability approach will require one to consider every possible capability in the individual's overall capability set, with each related functioning. But this is clearly not the case. The qualitative work that Verkerk *et al.* mention will be conducted to develop a

list, or subset, of valuable capabilities or functionings relevant to the particular research question; a natural and integral step in operationalising the approach, regardless of the discipline or context in which the approach is being used. In the case of health this will mean listing capabilities from the overall capability set on which a particular health state or health intervention will impact.

In a commentary on the paper by Verkerk *et al.*, Hurley (2001) highlights a similarity between existing work in the field of health-related quality of life and the notion of functionings. Hurley states that Health status classification systems such as EQ-5D or SF-36 emphasize a person's level of functioning in important domains of health-related quality of life (Hurley 2001, p58). This has been acknowledged earlier in this chapter.

Despite this similarity, Hurley believes that there are some 'deep challenges' for translating Sen's ideas into practical measures of health-related quality of life. One of these challenges is informational constraints (as acknowledged in section 5.3). The other perceived challenge is that the capability approach includes a broader array of elements than is normally included in health-related quality of life studies. Once again, this does not have to be regarded as a challenge at all; if one chooses to focus only on health then the result will be a simple health-status classification system. If on the other hand, there is a desire to give more legitimacy to the final list of capabilities/functionings and to attempt to include all of the functionings that patients have reason to value then this is possible through the use of the capability approach.

If patients value wider quality of life then it may be appropriate to avoid artificially distinguishing health-related aspects of quality of life from the impact that health has on *overall* quality of life.

Cookson (2005) also suggests that the QALY approach is recognizable as an application of the capability approach since it pays close attention to functionings, through the use of survey based multi-attribute health status classification systems, and permits conceptions of value other than the traditional utilitarian ones of choice, desire-fulfilment and happiness. Cookson

reinterprets the QALY as a cardinal and interpersonally comparable index of the value of an individual's capability set in a given time period under uncertainty (Cookson 2005, p818). Here 'value' does not necessarily need to be interpreted in terms of any of the standard utilitarian senses of utility (p823).

Cookson's interpretation of the QALY differs from the usual interpretation as an index of health, whereby the QALY represents one important component of an individual's well-being – the health component. Cookson's QALY represents all of the individual's well-being. Standard health state valuation methods specify the health state, but not the wider capability set; hence such methods are said to value health-related quality of life. Under Cookson's reinterpretation of the QALY, the use of standard health state value elicitation methods to value capability sets requires an assumption that individuals factor in their own perception of their (unspecified) capability set and changes therein (Cookson 2005, p823). Cookson's Capability QALY represents the value of the capability set induced by specifying health states in a given time period. Put simply, valuation survey respondents are assumed to consider how their health will affect their freedom to choose non-health activities and states (Cookson 2005b).

It is recognised by Cookson that there is no mechanism for making sure that different individuals are considering the same list of functionings when formulating their answers to questions contained in surveys such as EQ-5D. Nor is there any mechanism for correcting potential misconceptions and errors in people's predictions about how changes in health status will influence their broader capability set. It would appear unreasonable to expect general/healthy respondents to have a ready understanding of how a particular health state would impact on their capability set. If we can discover from talking to patients themselves what impact their health state has on their capability set then it may seem more logical and consistent to value this explicitly defined impact, rather than relying on respondents (who are unlikely to have any direct or indirect experience of the health state) valuing their *perceived* impact.

In response to Cookson, Anand (2005) expresses doubts over the suitability of measures such as EQ-5D as the foundation of a capability QALY and suggests that these measures need to be extended. Anand is cautious about accepting Cookson's claim that the QALY can represent all of the individual's well-being (Ibid. p1286). Anand argues that: "to provide such a representation explicitly, a survey instrument would need to address the individual components of well-being opportunities and functionings... this suggests further development of QALY type measures would be needed..." (Anand 2005, p1286). Anand concludes that there is strong reason to believe that such development is feasible.

Cookson responds by arguing that current instruments (such as EQ-5D) are perhaps more fit for purpose than Anand suggests. The functionings which are valuable will depend on the decision context; Cookson argues that generic health state descriptive systems have been designed specifically to inform decisions about health care resource allocation and are likely to be more relevant than the functionings considered in the capability literature so far. A second reason is that dimensions of EQ-5D such as 'usual activities' already cover more than just health. The first case in favour of existing health measures is not a strong argument. The fact that the dimensions within EQ-5D are more relevant to healthcare in a developed economy than are some other functionings discussed within a separate discipline does not automatically mean that EQ-5D is fit for purpose. With regard to the second case, the inclusion of one broad dimension does not make the EQ-5D a particularly broad measure of health; in fact compared with all of the possible dimensions of health discussed in chapter two, the EQ-5D is a relatively narrow measure of total health.

To summarise, Cookson's reinterpretation of the QALY provides us with a situation in which respondents are left to value how they imagine a set of narrowly defined health states might impact on their broader capability set. Respondents are unlikely to have any real experience of the health states in question (direct or indirect), and there is no way of knowing that respondents are considering similar lists of functionings, or what misconceptions are underpinning their responses. An alternative would be to explore more

precisely how a particular health state impacts a sufferer's capability set and, by doing so, obtain and define a relevant list of the key impacts on wider quality of life (i.e. capabilities/functionings).

5.6 Summary

This section has introduced the Capability Approach of Amartya Sen, defined the key terms associated with the approach, and has attempted to seek out guidance from the development literature on how the approach can be operationalised. The reader could be forgiven for feeling that more questions have been raised than have been answered. Sen has very deliberately refrained from 'filling in all of the detail' and prescribing how his theory should be used because he wants the Capability Approach to remain a general approach that can be used for a great variety of uses and in a great variety of contexts (Sen 1999; Qizilbash 2002). Others (Alkire & Robeyns) express a concern that unless "user-friendly operational procedures arise and spread, then interest in the capability approach is vulnerable to subversion by misinterpretation" (Alkire 2005, p116).

Intelligent and enthusiastic discussion has surrounded the capability approach but much of this discussion remains abstract and deeply theoretical. Chapter six will contain a review of the small number of cases where the capability approach has been applied in practice. Only when an approach is put to the test can its full range of weaknesses and advantages become apparent.

The capability approach has begun to receive attention within the discipline of health economics and it has been noted that the approach shares some key similarities with the QALY approach. Although some health economists have suggested the capability approach as a more general framework of quality of life this very characteristic of the approach has also been seen as a challenge. There seems to be a misconception that to use the Capability Approach to assess the impact of health one would need to look at every possible aspect of quality of life, whereas one would only need to consider the specific impact that a health state has on the overall capability set. In other words, a subset of

relevant and valuable capabilities would need to be developed; not every capability in the individual's overall capability set will be relevant or indeed valuable enough to justify inclusion in the sub-set of capabilities selected to address the research question.

It may be argued that this subset of capabilities is formulated and considered by the respondent anyway when valuing the health states defined within generic health-related quality of life measures such as EQ-5D. At this stage at least, this remains an assumption. Furthermore, if respondents do consider the impact of a health state on the overall capability set when valuing the health state, questions can be raised regarding the accuracy and consistency of the perceptions underpinning the valuation.

It has been argued in this chapter that it will be desirable to adopt some means of gauging the public's consensus as to the ranking and possibly the relative importance of the capabilities. Sen draws our attention to the disadvantages of using utility as the object of value. It should be noted that Sen uses the term utility as a catch-all term for some concept of well-being measured in terms of happiness, desire or choice. At this stage in the discussion no distinction has been made between cases where there are probabilities involved in the measurement task and those where there is not.

The distinction which is important is that between utility as the object of value and the use of 'utility' for the valuation of well-being, defined in some other way. In an effort to avoid confusion, techniques relying on expected happiness, desire or choice to value some concept of well-being defined in an alternative form (i.e. Capability) will from now on be termed 'preference-based' measures. Sen does seem to accept that preference will, in many cases, be the best indication that we have about the value of an alternative object and therefore it would seem that the weighting of Capabilities based on preference is acceptable, given the theoretical foundations of the Capability Approach. What is more, it has been argued that for practical reasons some form of valuation is indeed necessary.

CHAPTER 6: Systematic Review: Practical Applications of the Capability Approach

6.1 Introduction

It was noted in chapter five that there are a number of possible uses for the Capability Approach (e.g. criticising or defending moral norms, assessment of well-being and wider quality of life, and as a means of assessing inequality) and that there are a number of concepts of advantage on which one could focus (i.e. well-being achievement or freedom, agency achievement or freedom). In this chapter consideration will be restricted to examples of where the approach has been used to assess well-being and wider quality of life. In chapter five, a number of steps or processes were identified as being important when operationalising the Capability Approach, these were:

1. A decision as to which of those concepts of advantage listed in chapter five (Figure 5.2) to focus on.
2. A decision as to whether to focus on functionings or capabilities.
3. Identifying the objects of value.
4. Judging achievement in terms of a particular functioning/ capability.
5. Establishing the value of these objects.

Steps (1) and (2) simply involve a choice. If the purpose of a study is, for example, to assess standard of living then well-being achievement would be the relevant concept of advantage, and as a basic approach it would be acceptable to focus on functionings rather than capabilities. If the objective is to assess wider quality of life then the concept of advantage will need to be either well-being freedom or agency freedom; the focus will then need to be on capabilities. It was noted in chapter five, however, that informational problems or measurement constraints may require that functionings are used rather than capabilities. Question (3) requires the formulation and justification of a list of functionings/capabilities. In chapter five it was seen that there are arguments both for and against weighting capabilities, and so (5) will require a decision over whether to weight the selected capabilities; if the decision is yes, then a

weighting system will need to be chosen and used. In this thesis the fourth step has been explicitly introduced to encourage greater consideration of how to assess the achievement of functionings.

It was noted that the arbitrary setting of a 'pass mark' will determine, to some degree, whether an individual can be said to achieve a particular functioning. It was argued though that it may be beneficial to consider a person's *level* of achievement along the full spectrum, rather than simply using 'pass' and 'fail'.

One could reasonably expect that practical applications of the Capability Approach will follow at least the first three of these steps, although the choices made in response to questions (1) and (2) may be inferred rather than explicitly stated. How authors have approached each of these steps will be of interest.

A search of the literature was conducted to find cases/studies where the Capability Approach has been operationalised. Articles were reviewed in order to discover if, when the approach has been operationalised, each of the five steps above have been addressed in turn and how researchers have dealt with each. The review was specifically intended to discover:

1. How many cases there are of the approach being operationalised
2. In which contexts the approach has been operationalised (e.g. health, disability, international comparisons of living standards, assessment of poverty, etc)
3. If authors have explicitly stated which concept of advantage is being assessed
4. If researchers have focused on functionings or capabilities
5. How relevant functionings/capabilities have been selected
6. If, and how, authors have sought to judge the achievement of these functionings/capabilities
7. If, and how, researchers have attempted to value/attach weights to the functionings/capabilities

6.2 Search Methods

A search for articles discussing cases where the Capability Approach has been operationalised was conducted in the spring of 2007 using the same six on-line literature databases as used to find the economic evaluations (see chapter four). Although the capability approach has been developed in the discipline of development economics it has been used in other, related disciplines, such as health economics. The author is aware of at least one group (Grewal, Lewis et al. 2006) who have published articles in the health economics literature, in which the capability approach has been discussed. It was therefore felt that it was important to search both economic and social science databases and those specializing in medicine, nursing and healthcare systems. A list of search terms and summary of search results can be found in Table 6.3.

As before, articles were initially assessed based on title and abstract; articles were excluded at this stage based on the exclusion criteria set out in table 6.1. Articles not excluded were saved and read in full and a final collection of articles were selected from these according to the inclusion criteria, as summarised in Table 6.2. A large number of articles were excluded at the first screening because including the word 'capability' among the search terms results in articles being retrieved which relate to business management or electronics/computing. Other search terms were added, such as "Sen", "well-being", and "quality of life" to narrow the results and this significantly reduced the number of 'hits'. Where early searches returned many thousands of hits they were abandoned and the search narrowed as just described.

Additional information on the Human Development Index was found on the website of the United Nations Development Programme (<http://hdr.undp.org/reports/global/2003/faq.html>). In addition to the search outlined above, correspondence with those working in this area resulted in three papers being supplied directly by the authors. The paper by Coast, Flynn *et al.* (2006) is a conference paper, as is that of Lorimer and Lorgelly (2007), a working paper is also included by Anand, Hunter, *et al.* (2007). The initial process is summarised in Figure 6.1, although it is important to note that some

further papers (including two of those supplied directly by authors) have been added since the search as they have become available.

Table 6.1: Exclusion criteria (Capability Approach)

- Non-English language
- Articles not relating to the Capability approach
- Articles which discussed the capability approach only at an abstract/theoretical level (No application)

Table 6.2: Inclusion Criteria (Capability Approach)

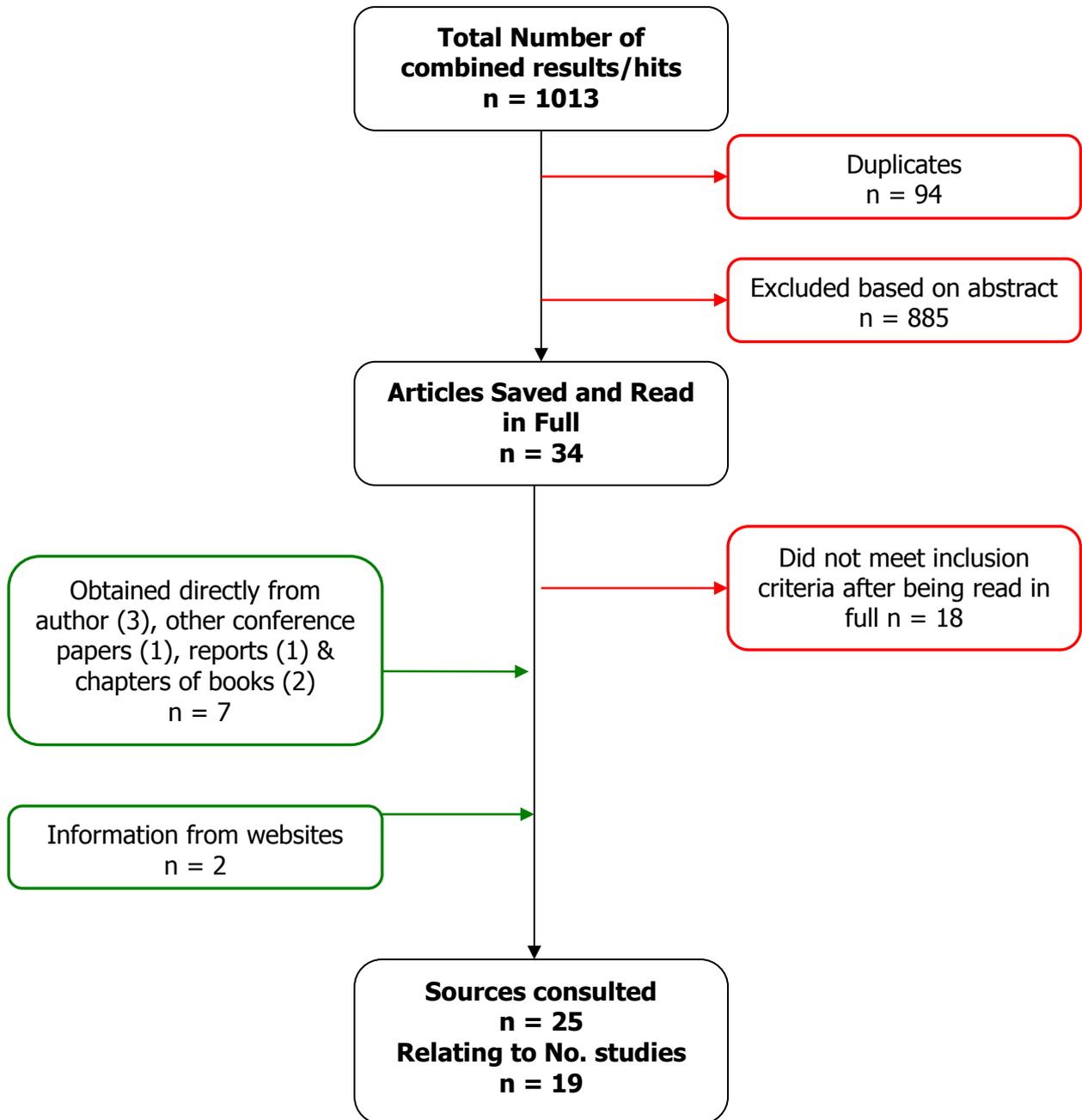
- English Language
- Articles detailing empirical work involving capabilities/functionings
- Articles commenting on empirical studies/ providing a critique of applications of the capability approach

Table 6.3 Search Terms and Summary of Search Results: Capability

Date Searched	First Search Term		Second Search Term		Individual Results						Combined Results	Duplicated Results	Initially Excluded
	By	Term	By	Term	EbscoH	EconLit	MEDLINE	PubMed	Science Direct	Web of Know'			
Mar 07	Subject	capability approach	-	-	0	0	0	0	12	(6834)	12*	-	11*
Mar 07	Subject	capability approach	All Fields	Sen	0	0	0	0	2	44	46	1	34
Mar 07	subject	functionings	subject	capability	0	0	0	0	1	8	9	5	3
Apr 07	subject	capability	subject	well-being	0	0	0	0	2	(9909)	2*	1*	1*
Apr 07	Title	capability	Subject	well-being	0	0	0	0	0	378	378	19	369
Apr 07	Title	well-being	Subject	capability	0	0	0	0	0	95	95	1	88
Apr 07	Title	capability	Subject	quality of life	7	0	5	0	1	17	30	6	22
Apr 07	Title	selecting capabilities	-	-	1	0	0	1	0	3	5	3	2
Apr 07	Title	health	Subject	capability	1	1	0	0	4	221	227	4	222
Apr 07	Title	capability approach	Subject	Operational*	2	2	0	0	1	2	7	1	5
Apr 07	Title	capability approach	Subject	Application	1	5	1	1	4	7	19	5	11
May 07	Title	physical quality of life index	-	-	0	5	2	0	4	16	27	6	16
May 07	Title	human development index	-	-	14	59	7	14	11	46	151	42	98
May 07	Title	functionings	Subject	well-being	0	4	0	0	0	1	5	0	3
Total											1013	94	885

* Where a search of one particular database gave many thousands of hits the search was abandoned and a new search conducted with additional search terms

Figure 6.1 Summary of Literature Search: Capability



6.3 Search Results

Sixteen cases where the approach has been operationalised were found through the initial literature search, a further three have been added since then; all of the cases are listed in Figure 6.1 and in Table 6.4 below. Table 6.4 indicates whether each study has used functionalities or capabilities, whether the concept of advantage is clearly stated, how the capabilities/functionings have

been obtained, if the authors define what they mean by the achievement of a particular functioning/capability, and if and how functionings/capabilities are weighted. Where the authors have not explicitly stated which concept of advantage is being assessed this has been entered in brackets.

Table 6.4: A Summary of Practical Applications of the Capability Approach

Application/ study	Focus (Context)	Concept of Advantage: Stated?	Functionings or Capabilities?	How do the authors develop a List of Capabilities?	Do the authors define & Assess Achievement?	Do the authors weight/ Value Capabilities?
Physical Quality of Life Index (PQLI) (Morris 1978)	International comparison of levels of economic development & standards of living	Not explicitly stated (Well-being achievement)	Achieved functionings used to indicate achievement in capability	Proxy variables selected by the author. Selected as non-ethnocentric variables for which data is commonly available.	Natural units (e.g. life expectancy in years) are transformed into index values between 1 (poorest performance for a given base year) and 100 (best possible performance)	Components (functionings) given equal weight (arithmetic mean used)
Human Development Index (HDI) (Launched 1990)	Measurement of standards of living in each country	Well-being (achievement)	Achieved functionings used to indicate achievement in capability	Proxy variables selected by authors	Separate index calculated for each dimension.	Equal weight (i.e. 1/3) given to each dimension.
Klasen (1997 & 2000)	Analyses poverty and inequality in South Africa based on data from a household survey	Well-being (achievement)	Functionings	Dimensions selected from 1993 SALDRU Survey (South Africa)	Ranking system	Equal weights (the average is used)
Sen, A. (1999, pp.46-51)	Example of the disparity between GNP and achievement in the capabilities of survival and education	Well-being (Achievement)	Proxy variables (i.e. achieved functionings) used to indicate level of achievement in capabilities of survival and education	Variables selected for which data is collected and included in <i>World Development Reports</i> 1983-1984.	In natural units (e.g. life expectancy, or higher education ratio)	No weighting (simple discussion & comparison)
Sen, A. (1999, pp. 52-69)	Judging the well-being of men vis-à-vis women in India	Well-being (achievement)	Functionings	Chosen by author	In natural units (e.g. rates of malnutrition)	No weighting (simple discussion & comparison)

Table 6.4 continued

Application/ study	Focus (Context)	Concept of Advantage: Stated?	Functionings or Capabilities?	How do the authors develop a List of Capabilities?	Do the authors define & Assess Achievement?	Do the authors weight/ Value Capabilities?
Mazumdar, K. (1999)	Analyses the achievement of quality of life in 68 developing countries using an extended PQLI	Well-being (achievement)	Functionings	Authors select two new functionings, which are used to extend those incorporated in PQLI	As for PQLI	As for PQLI
Alkire (2002)	Assessment of three Oxfam funded projects aimed at improving quality of life in rural areas of Pakistan	Not explicitly stated (Well-being achievement)	Functionings	Participatory approach	Two separate ranking systems developed	Capabilities & achieved functionings given equal weight
Phipps, S. (2002)	well-being of young Canadian children in international perspective	Well-being (achievement)	Functionings	Authors select a set of functionings for which data is available	Response frequency	Equal weights attached
Clark (2005)	Perceptions of development among South Africa's poor	Well-being (achievement)	Commodities/ activities & functionings	Participatory approach	-	-
Anand, P, Hunter, G, et al. (2005 & 2007)	The construction of a new survey instrument for the measurement of individual capabilities & exploration of the link between capability and life satisfaction	well-being achievement	Functionings	Use of Nussbaum's list of Central Human Capabilities	Through the use of questions from the British Household Panel Survey	No
Anand & van Hees (2006)	To explore relations to life-satisfaction (happiness) in a random sample of English voters.	Well-being & Agency Freedom	Capabilities	Selected by authors	Respondents rank satisfaction with each individual capability	-
Grewal et al (2006)	Development of a generic quality of life measure for older people in the UK	Not Explicitly stated (Well-being & Agency Freedom)	Capabilities	Participatory approach	-	-
Coast, J. et al. (2006, 2007, 2008, 2008b)	Discrete choice modelling used to obtain values for attributes developed by Grewal, Coast et al. (2006)	Not Explicitly stated (Well-being & Agency Freedom)	Capabilities	As set out by Grewal et al. (2006)	Four levels of achievement specified by authors	Using discrete choice experimentation

Table 6.4 Continued

Application/ Study	Focus (context)	Context of Advantage: Stated?	Functionings or Capabilities?	How do the authors develop a list of capabilities?	Do the authors define & assess achievement?	Do the authors weight/ value capabilities?
Biggeri et al (2006)	Children's well-being	Not explicitly stated	Capabilities	Participatory approach	-	-
Chowdhury, S. & Squire, L. (2006)	Alternative weighting system for HDI	Well-being (achievement)	Functionings	Use of HDI	As for HDI	Opinion survey with panel of experts
Wigley, S. & Akkoyunlu-Wigley, A (2006)	Gauging the value of education in 35 developing countries	Not explicitly stated (well-being achievement)	Functionings	Authors select life expectancy	Years of life	n/a
Lorimer, K., Lorgelly, P. et al (2007) (2008)	Public Health in the UK	(Well-being achievement)	Functionings	Focus is on refining the list of functionings used by Anand, Hunter, et al. (2005 & 2007)	Through use of questions from British Household Panel Survey	Not at present
Frediani, A. (2007)	Evaluation of a squatter upgrading programme in Salvador de Bahia	Not explicitly stated (well-being achievement)	Functionings	Focus groups with young adults from the squatter settlement	Not clear	Not Clear
Schischka, J. et al. (2008)	Identification of capabilities which have changed as a result of two poverty alleviation projects	(Agency Freedom)	Capabilities	Focus groups with project participants	No	No

6.4 Discussion

As noted above, 19 cases were identified in which the Capability Approach has been operationalised. It was expected that the studies would have addressed, in turn, at least the first three of the five steps outlined at the beginning of this chapter. In fact, the review reveals that researchers often focus on one or two of these five steps, without detailing or justifying how they reached that step and without intending to progress from it. There are numerous examples of valuable functionings and capabilities provided in the literature, ranging from educational attainment to being able to live with concern for the world of nature. There are few cases, however, in which a context specific list of capabilities has been justified or in which lists have been developed using a participatory approach involving the community to which the list is supposed to be relevant. This may be difficult when it comes to assessing and comparing poverty and development for entire countries. Many studies conducted at a macro level simply use whatever data is available (Morris 1978). But, there are examples where the Capability Approach has been used to evaluate small-scale projects where lists have been developed through qualitative methods. In one case (Clark 2005), the sole objective of the study was to develop a list of capabilities through participatory methods. There is only one case where capabilities have not simply been given equal weights (Coast, Flynn et al. 2005), although there are some interesting cases of where achievement in terms of capability has been defined and judged.

In answer to the third question, it was found that very few authors state explicitly which concept of advantage they were seeking to assess. Those studies in which functionings were assessed and those in which capability was the focus can be identified from Table 6.4. Many of the studies (10) restricted their focus to functionings; in five studies the focus was on capability. Clark (2005) appears to work with capabilities and functionings. In several studies it appears that achievement in terms of capabilities is inferred from a number of available proxy variables. A more detailed response is needed for questions 2, 5, 6 and 7. Starting with question two, section 6.4.1 will outline the different contexts in which the approach has been used, giving further detail about the individual

studies. In section 6.4.2 the processes through which capabilities and functionings have been identified/selected will be discussed. Section 6.4.3 will address questions six and seven.

6.4.1 The different Contexts in Which Capability has been Applied

Eight studies assess well-being at a national level, most conducted within the context of overseas development. Two of these studies were conducted with the principal aim of demonstrating the disparity between well-being assessed in terms of national income and capability based indications of national well-being. One study was used to highlight the differences in well-being between men and women in India. In three cases the Capability Approach is used to make international comparisons of well-being. In one case the well-being of Canadian children is assessed. Wigley *et al.* (2006) use capability to establish the value of education within developing countries. The remaining studies either evaluate small scale projects or use household survey data to assess achieved functioning.

Wigley and Akkoyunlu-Wigley (2006) use panel data analysis for 35 developing countries and compare the health functionings (as indicated by life expectancy) which are achieved by income growth generated by educational attainment, with the total health functionings which are achieved by educational attainment. The authors found that educational attainment has a significant effect on life expectancy independently of its effect by way of income growth. It is concluded that income underestimates the health functionings achieved by educational attainment. Well-being achievement is the concept of advantage considered in this case, although this is not explicitly stated in the article.

Sen (1999) demonstrates the disparity between well-being assessed in terms of GNP and well-being assessed through a series of functionings for five developing countries. Once again, although not stated by the author, well-being achievement is the concept of advantage being considered; a small number of functionings are simply selected by the author for which data is available.

In the study by Phipps (2002) the Capability Approach is used to compare levels of well-being in three developed countries (Canada, Norway and the United States). Phipps uses the approach to compare the well-being of young children.

Possibly the two best known measures of well-being in which capabilities are weighted and developed into an index are the United Nations Development Programme's (UNDP) Human Development Index (HDI), and the Physical Quality of Life Index (PQLI). Morris's Physical Quality of Life Index was the first well known index to focus entirely on direct measures of quality of life, rather than on individual or national income. The PQLI includes life expectancy, infant mortality and literacy. The reason for this focus on direct measures of quality of life was a concern that different individuals need different amounts of income to live a life that can be considered valuable; it is widely considered that this basic insight is also relevant at a national level, so that countries with similar GNP per capita could still vary in terms of achievement of quality of life (Qizilbash 2002, p4); this has already been seen in this chapter through the work of Wigley and Akkoyunlu-Wigley (2006) and Sen (1999). The UNDP's Human Development Index was developed in a similar way to the PQLI.

Alkire (2002) has used the capability approach to assess three Oxfam funded projects aimed at improving quality of life in rural areas of Pakistan. Alkire identified human impacts, or changes in the capability set of participants, from each of the projects using participatory methods. The impact of projects on a particular capability was ranked on a scale of zero to five, relative to what was deemed possible in the situation. Negative scores were also allowed.

Very much following the lead of Alkire, Schischka *et al.* (2008) used focus groups to explore the changes in the capabilities of participants in two very different poverty alleviation programmes. The first case study is a programme based around community gardens in a low income neighbourhood in the city of Christchurch in New Zealand. The second was a self-help project for village women in Samoa. The Community Garden Project was operated over three locations, with each location catering for a distinctly different target group

(retired men; people with learning difficulties; those accessing the project via the Young Men's Christian Association).

The second case study involved a programme operated by the Women in Business Foundation in the Island nation of Samoa. The projects here involved small-scale technologies (such as a process for producing coconut oil for export or weaving mats for sale to tourists) appropriate for village women to provide much of their own income needs.

Work by Grewal *et al.* (2006) was conducted as part of a wider programme (Investigating Choice Experiments for the Preferences of Older People) and fits directly alongside the work of Coast *et al.* (2006). This ongoing work is significant because it is being conducted in the UK by a team of health economists. The aim of the study was to determine attributes for a new index clearly focussing on quality of life for older people rather than health, which was deemed to be an influence on quality of life.

Clark (2005) conducted fieldwork in 1998 among South Africa's poor to investigate perceptions of development. Biggeri *et al.* (2006) conducted a research project that allowed children to define their capabilities as the basis of a strategy for understanding the relevant dimensions of children's well-being.

Finally, Anand and van Hees (2006) developed a questionnaire to explore life satisfaction of individuals in the UK.

6.4.2 Obtaining lists of capabilities

Wigley and Akkoyunlu-Wigley (2006) in their study looking at the effects of education on health, report having chosen life expectancy as a proxy for health functioning and average years of schooling to indicate educational attainment. There is an acknowledgement that compared to literacy, educational attainment provides an indirect indicator of the capability for educational functioning. The PQLI considers well-being achievement and uses a number of functionings deemed by Morris to be non-ethnocentric and which were selected, once again,

largely on the basis of the availability of data. The fact that income is included in the HDI makes it distinctly different from the PQLI, which only includes direct measures of quality of life.

Mazumdar (1999) takes the view that quality of life should be measured on the basis of as large a number of attributes as is relevant and feasible and therefore attempts to widen the scope of the PQLI by incorporating two more variables (percentage of total population living in urban areas and per capita calorie supply as percentage of requirement).

The method used by Alkire (2002) to obtain a list of 'impacts' from the Oxfam projects was to explain the general intent of the assessment exercise (to think about the full range of impacts of an activity, good and bad, anticipated and unanticipated) and to begin by asking an open question: what valuable and negative impacts have you noticed? As people described them the impacts would be further explored (Alkire 2002, p225). Having identified a wide set of impacts, the next stage was to separate the relatively trivial impacts from the central ones. This was done in two ways; one method was to arrange items into clusters and ask groups of respondents to rank them from strongest to weakest, with items being allowed the same ranking. The second method was simply to ask participants to identify their top three categories of impact.

Schischka *et al.* (2008) conducted focus groups with the two communities involved in the Community Garden Project in New Zealand and the Women in Business projects in Samoa. Starting with the Community Garden Project, focus groups were held on location at the gardens. A discussion guide was used for the focus groups, each of which lasted for between 30 and 60 minutes. All participants in the project were invited to attend the focus groups, but it is unclear how many participants there were in each group or how many groups were held. Participants were asked what their expectations were upon joining the project, and what new choices and opportunities they have had since their involvement. New capabilities found in the Christchurch case study are summarised in Figure 6.2.

Figure 6.2 New Capabilities in the Christchurch Case Study

1. The ability to learn and apply more gardening skills
2. The ability to have social contact and be part of the community
3. The ability to lead healthy lives
4. The ability to have increased self-confidence and status
5. The ability to earn future income

For the Women in Business project, the research team once again conducted focus groups involving participants on location in the villages. Thirteen focus groups were conducted during January 2003. New capabilities associated with the second case study are reproduced in Figure 6.3.

Figure 6.3: New Capabilities in the Samoa Case Study

1. The ability to generate cash income from local sources
2. The ability to support the family
3. The ability to make handicrafts for sale
4. The ability to revive local crafts
5. The ability to contribute to the local church and community

What is interesting about both case studies, although perhaps not surprising, is that participants were able to reflect on the way the programme had helped them to discover capabilities they already had but which they had not previously realised could be valuable in creating self-help opportunities. The case studies investigated by Alkire and Schischka seem to have been more about supporting, guiding and encouraging individuals to channel their capability into achieved functionings, which would be unlikely to have been observed otherwise. The work by Alkire, and later that of Schischka *et al.* resulted in lists which include non-monetary gains which were specific to particular projects. As will be outlined in the following section, the work of Alkire further obtained through focus groups qualitative gradings for particular capabilities.

Recall that the study by Grewal *et al.* involved identifying capabilities important to older people. In-depth interviews were conducted with 40 informants aged 65 and over in private households to explore their views about what is important to them in terms of quality of life. Interviews were opened with broad questioning about what was important to respondents, what they enjoyed, got pleasure from or valued in their lives. The intention was to allow the older people to articulate their thoughts about quality of life in their own language and in their own way. Researchers then used responsive questioning to probe for underlying attributes of quality of life. A topic guide was used to ensure systematic coverage of issues across interviews. The data was analysed using framework analysis.

Five attributes were identified: attachment, role, enjoyment, security, and control. The research team noted that the quality of informants' lives was limited by the loss of ability to pursue these attributes, which led them to interpret the five attributes as a set of functionings (Grewal, Lewis *et al.* 2006). It was therefore concluded that further development of the measure should focus on an index of capability, rather than preference-based utility.

A second stage of qualitative work was conducted in order to refine the language used to express the five attributes, involving semi-structured interviews with 18 of the informants from the previous work (Coast, Flynn *et al.* 2006). The final language used was:

- “Love and Friendship” (attachment)
- “Doing things that make you feel valued” (role)
- “Enjoyment and pleasure” (enjoyment)
- “Thinking about the future without concern” (security)
- “Being independent” (control)

In the study by Biggeri *et al.* (2006) children aged between 11 and 17 and attending the ‘Children’s World Congress on Child Labour’ in Florence, May 2004, were invited to share their opinions on the most relevant issues related to their childhood and adolescence. The children assembled in Florence were

from different regions, countries, organisations, and cultures, although a common factor was that they were mainly from poor backgrounds and that many were former child labourers.

The study follows a growing concern among international agencies that children have a right to express their opinions freely and policymakers and agencies should take account of this opinion in any matter affecting the child (Biggeri, Libanora et al. 2006). It is argued that there are difficulties in the assumption that (outsider) adults should be able to determine the best interests of (insider) children. One could argue that there are similarities here with the Doctor and patient relationship; in chapter three it was reported how Birch and Donaldson (2003) stress the importance of taking into account a patient's opportunity cost when deciding what is in the patient's best interests.

Although the study by Biggeri *et al.* focuses on the capabilities of children, and in a context rather different from that in which the Capability Approach will be used in this thesis, the study is nonetheless deemed relevant because it is one of such a small number of studies in which a list of capabilities has been obtained through participatory methods.

As a starting point the research team used Biggeri's *ad hoc* list of 14 children's capabilities (Biggeri, Libanora et al. 2006), although a participatory approach was actually used in order to identify a final list of capabilities for children. The categories in Biggeri's initial list are as follows:

1. Life and physical health
2. Love and care
3. Mental well-being
4. Bodily integrity and safety
5. Social relations
6. Participation
7. Education
8. Freedom from economic and non-economic exploitation
9. Shelter and environment
10. Leisure activities
11. Respect
12. Religion and identity
13. Time-autonomy
14. Mobility

A combination of a questionnaire, focus group discussions and in-depth interviews were used. A particular capability was included on the final list if it appeared on Biggeri's initial list and a majority of children considered it as important or very important, or if at least one child identified it without any interference. All of the capabilities on Biggeri's list were indicated to be important, although some capabilities were more frequently identified than others (education, love and care, leisure activities, and life and physical health).

The only new capability to be added to the list following the qualitative and quantitative work was 'paid work'; this is likely to reflect the children's background. It is interesting that despite some children listing 'paid work' as a capability, none of the children actually thought that paid work was an important opportunity a child *should* have during their life. This can be taken as proof that human perception and desires are influenced by personal and social history. It appears children recognised the importance that child work can have in contributing to and sustaining the family economy in times of need and for this reason listed child work as a capability (Biggeri *et al* 2006, p76).

As well as the final list of capabilities, two important findings were revealed through the study. First, that irrespective of their specific background, among the children in the sample there was a common view on core capabilities. The second observation was that the relevance of certain capabilities varied according to age.

The study by Biggeri *et al.* is also significant because it requires that those who advocate the capability approach give further consideration to the issue of adaptation. Sen's discussion on deprived individuals lowering their expectations and taking pleasure in the 'simple things in life' is well rehearsed. It is also accepted that people in desperate poverty will pay more attention to basic capabilities, such as shelter and access to clean water, than would people in more fortunate circumstances for whom these capabilities are no less important, but which tend to be taken for granted. The inclusion of 'child work' by the children in this study suggests that individuals living with great hardship

may not only focus on more basic capabilities and perhaps exclude capabilities which they cannot hope to possess, but may also report some capabilities which are simply different or unique.

The question for academics is how to deal with these capabilities. Is it acceptable to dismiss capabilities such as 'child work' on the grounds that they are a product of adaptation? The central theme of the capability approach is freedom, but it is not clear whether we should restrict the ability of a family to receive additional income in times of need in order to defend the freedom of a child from economic exploitation. This is, however, an issue which falls outside the scope of this thesis.

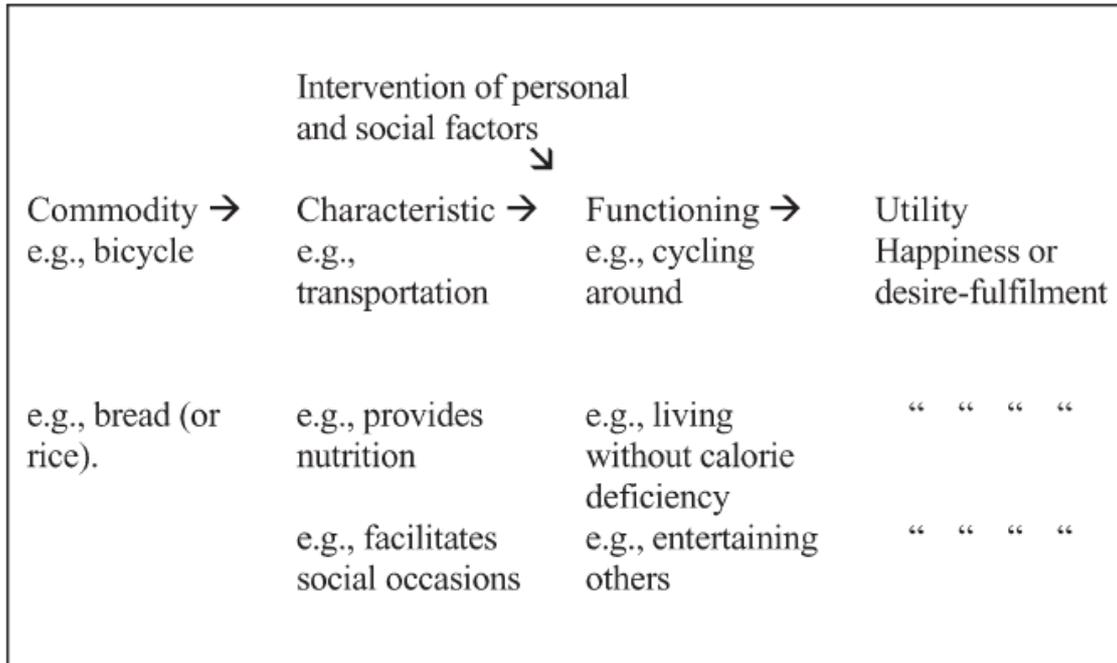
Clark (2005) conducted fieldwork in 1998 among South Africa's poor to investigate perceptions of development. One hundred and fifty seven in-depth interviews were conducted with respondents from a rural village and an urban township to discover how the poor view a 'good' quality of life. Interviews were based around a questionnaire, which contained open and closed questions. Clark illustrates the relationship between commodities/ activities and functionings reported as valuable by South Africa's poor, some of which are reproduced in Table 6.5.

Table 6.5: Some Examples of Commodities/Activities, Their Characteristics and the Achievements They Help to Facilitate

Commodity/ activity	Characteristics	Functionings
Coca-Cola (and other soft drinks)	<ul style="list-style-type: none"> • Low in protein & fat • High in carbohydrates/sugar • Provides for liquid needs • Clean, safe & hygienic • Medicinal properties • Yields utility • Fashionable product • Social drink 	<ul style="list-style-type: none"> • Some nourishment • Provides energy, but rots teeth • Meets liquid needs, quenches thirst • Avoid ill-health • Relieves stomach ache, wind, diarrhoea • Provides a 'simple moment of pleasure' • Feelings of being 'transported' to the West • Being fashionable • Facilitates social activities
Television & Cinema	<ul style="list-style-type: none"> • Education • Yields Utility • Recreational • Social Activity 	<ul style="list-style-type: none"> • Learn new things • Satisfy curiosity • Misleading/undermines social values • Pleasure & entertainment • Escapism • Facilitates daydreaming & imagination • Avoid boredom & mischief • Facilitates social life
Advertising	<ul style="list-style-type: none"> • Provides Information • Promotes economic activity • Yields utility 	<ul style="list-style-type: none"> • Facilitates buying products • Promotes choice • Facilitates economic activity • Pleasure, entertainment & amusement • Dream of future

It is suggested by Clark that the examples of capabilities and functionings currently provided in the literature, by Sen and Nussbaum in particular, are both conveniently distinct from utility and restricted to things which the writer believes will unequivocally contribute to a healthy, active and rewarding life. Some of the commodities, characteristics of commodities and related functionings contained in Table 6.5 challenge Sen and Nussbaum's perception of what constitutes a good life. Reflecting on results from the fieldwork, Clark also concludes that the boundaries between functionings and utility are less clear-cut than is apparent from the literature on capabilities (see Figure 6.4 for an illustration of this relationship).

Figure 6.4: The sequence from commodities to utility:



Source: Clark (2005, p1344)

The point can be made that coca-Cola is not a particularly good source of nutrition, but that the product is hugely successful, even in some of the World's poorest countries (Clark 2005). It is true that respondents valued Coca-Cola as a 'healthy, clean drink', but they also rated enjoyment well above other motives for drinking Coca-Cola. In short, the drink provides a 'simple moment of pleasure' and is regarded as a superior first world product.

Sen's capability framework also has to deal with activities like watching television and visiting the cinema. Some of the poorest households frequently forgo many of life's necessities in order to acquire a television or radio. Clark (2005) found that respondents were keen to emphasise the importance of watching television to relax or to take their mind off other worries. Cinema trips and television programmes were also regarded as important sources of pleasure and entertainment and provided important forms of 'escapism', as well as helping to pass the time. While on face value many of the functionings identified by Clark would sit awkwardly in Sen or Nussbaum's lists, satisfying one's curiosity and sharing a social activity with friends and family do fit much better with the type of examples given by these authors.

It is clear that experiencing happiness is an important aspect of well-being that features prominently in most activities. The concept of utility, Clark notes, could be “stretched to consider psychological functionings as diverse as feeling relaxed, achieving peace of mind, day dreaming and experiencing pride” (Clark 2005, p1357). Sen does in fact acknowledge the value of utility; he refers to being happy as a “momentous functioning” that has “importance of its own” (Sen 1984, cited in Clark 2005, p1357). Clark is a supporter of Sen’s capability framework, and does believe that the approach benefits from a wider information base than standard utilitarianism, which effectively reduces well-being to a single category (utility). Clark concludes that the capability approach can handle complex notions of utility by simply treating different mental states as examples of valuable functionings (Ibid, p1360).

Anand and van Hees developed a questionnaire which was divided in to two main sections, satisfaction with own capabilities and perception of others’ capabilities. The dimensions chosen by Anand and van Hees (2006) are: achieving things, health, intellectual stimulation, satisfying social relations, pleasant environments, and the development of social projects. An example of a question from the first section of the questionnaire is as follows: (Anand and van Hees 2006, p280)

I feel the scope to seek happiness in my life is:

Very Good (1) ... Neutral (4) ... Very Inadequate (7)

The questionnaire was mailed out to 1,323 recipients in England, with just over 20% of recipients returning useable responses.

In separate, but related work Anand, Hunter *et al.* (2005) have identified questions contained within the British Household Panel Survey¹⁰ which the authors see as relating to the central capabilities listed by Martha Nussbaum (Nussbaum 2003).

¹⁰ <http://www.statistics.gov.uk/STATBASE/Source.asp?vlnk=1308&More=Y>

Nussbaum's list of ten central human capabilities are: (Nussbaum 2003)

1. Life: Being able to live to the end of a human life of normal length
2. Bodily Health: Being able to have good health
3. Bodily Integrity: Being able to move freely from place to place
4. Senses, Imagination, and thought: Being able to use the senses, to imagine, think, and reason
5. Emotions: Being able to have attachments to things and people outside ourselves
6. Practical Reason: Being able to form a conception of the good and to engage in critical reflection about the planning of one's life
7. Affiliation: Being able to live with and toward others, to recognise and show concern for other human beings; having the social bases of self-respect and non-humiliation
8. Other Species: Being able to live with concern for and in relation to animals, plants, and the world of nature
9. Play: Being able to laugh, to play, to enjoy recreational activities
10. Control Over One's Environment: Political and material

Some questions from the British Household Panel Survey (BHPS), identified by Anand as relating to capabilities from Nussbaum's list are contained in Table 6.6.

Table 6.6: Nussbaum's Capabilities and selected British Household Panel Survey Questions (Anand 2005)

Nussbaum's Capability	BHPS Wording
Bodily Health	Does your health in any way limit your daily activities compared to most people of your own age?
Bodily Integrity	Do you normally have access to a car or van that you can use whenever you want to?
Senses imagination and thought	Highest educational level
Emotions	Have you recently been able to concentrate on whatever you're doing?
Practical reason	Have you recently felt capable of making decisions about things?
Affiliation	Would you like to have friends or family for a drink or meal at least once a month, but must do without because you cannot afford it?
Control over one's environment	Does your health keep you from doing some types of work?

The BHPS also contains a number of other sections, so respondents are asked whether they are satisfied with their life overall, and then asked whether they

are satisfied with: their health, their flat or house, the income of their household, their partner, their job, their social life, the amount of leisure time they have, and with their use of leisure time. Anand *et al.* attempted to measure the link between capabilities and life satisfaction. Interpreting life-satisfaction as well-being, Anand *et al.* report finding 'strong evidence' that capabilities do influence well-being, although they also acknowledge that their findings may more accurately be said to measure functionings.

Anand and van Hees (2006) feel that their study is significant because they demonstrate that it is possible to design questions that distinguish between capabilities and achievements (functionings). Certainly the phrasing of the questions by Anand and van Hees appears novel; they ask respondents about the 'scope to do or achieve things' and about 'limitations of opportunities'. The phrasing of Anand and van Hees' questionnaire would appear a useful starting point for those planning similar work, although a degree of caution is needed. Although revisions to the wording were made after the questionnaire had been piloted, no information is provided by the authors to demonstrate that respondents fully understood the questions. It appears no attempt was made to assess how respondents interpreted the question and what considerations they took into account when answering. The authors therefore appear somewhat bold in their claim that they successfully designed a questionnaire that distinguished between capabilities and functionings.

In further work, Anand, Hunter, *et al.* (2007) collect new data using a measure of life satisfaction identical to that in the BHPS, and their set of capability questions. The focus of this study was on understanding which capabilities can be taken to be co-variates of life satisfaction. From this empirical work the authors developed a shorter list of capability indicator variables for which there is the strongest evidence of a statistical link to subjective well-being.

Lorimer, Lorgelly, *et al.* (2007) have also sought to further refine the list of 60 capability indicators selected by Anand, Hunter, *et al.*, with the aim of developing an index of capabilities that could be used to evaluate complex

public health interventions. This refinement involved two phases, in which a range of qualitative and quantitative methods were used.

Focus groups were involved in phase one, to develop an understanding of: comprehension of the question, retrieval of relevant beliefs, the forming of a judgement, and the selection of an appropriate response to each of the questions. The idea of the focus groups was to gain an understanding of this process of thought, rather than to simply gain answers to the questions. Also in phase one, analysis was conducted using previously collected quantitative data, specifically factor analysis was used, a technique which aims to simplify complex sets of data by attempting to describe correlations between variables. In other words, the technique indicates when questions may be redundant. A revised version of the questionnaire was then further assessed in interviews and in a pilot postal survey. Semi-structured interviews with members of the public were then conducted in an attempt to still further refine the questionnaire.

One interesting point that the authors note is that, while “most focus group participants had some comprehension of the difference between functionings and capabilities, whether the questions posed actually measured this difference is still unclear” (Lorimer, Lorgelly et al. 2007).

Phase two involved a thematic analysis of the interview data and a quantitative analysis of all completed questionnaires with the aim of identifying areas in which the questionnaire could be further reduced (Lorgelly, Lorimer et al. 2008). The questionnaire was then redesigned once again (to develop version 3) and further interviews and a postal survey followed.

6.4.3 Judging Achievement and Calculating an Index of Capability

The (proxy) variables used by Wigley and Akkoyunlu-Wigley (2006) were life expectancy, educational achievement and growth in income. Achievement in terms of health was measured in standard units, i.e. years. Because health was the only outcome considered in this case (education and income were inputs), no weighting was required.

In order to calculate a score for the PQLI, life expectancy, infant mortality and literacy are each placed on a fixed 1-100 scale. The worst performance for each component for a particular base year is set as 1. The best performance for the latest or current year is set as 100. It is assumed that all subsequent years will show movement above the worst performance in the base year. Having placed the three indicators on a common scale, the composite index is the result of an equally weighted summation (Morris 1978). Mazumdar's extended PQLI still converts the (now) five components into index values of between 1 and 100 and still gives each component equal weight.

The HDI incorporates three dimensions of human development: the ability to lead a long and healthy life, be knowledgeable, and have access to the resources needed for a decent standard of living. These three dimensions are aggregated into the HDI by simply averaging the three. Although the equal weighting of the three dimensions has been much criticised on the grounds that it is arbitrary, work by Chowdhury and Squire (2006) suggests that equal weights are consistent with the views of experts.

For each of the component indices a maximum and a minimum value are set as reference points. The life-expectancy index can be considered as an example. A maximum value for life-expectancy at birth is set at 85 years, with a minimum value of 25 years. So, for a particular country, with a life expectancy of x (years), one calculates $(x-25)/(85-25)$ (Gaertner and Yongsheng 2006, p312).

The minimum and maximum levels chosen for the three dimensions within the HDI have attracted controversy. Whilst the minimum and maximum levels for education (0% and 100% respectively) are natural choices for this dimension, the levels set for health and income are somewhat arbitrary (Qizilbash 2002, p5). It is also clear, when one considers how the three component indices of the HDI are calculated, that a change in, for example, the minimum or maximum levels of life expectancy will change the HDI value for all countries, and, if a fixed cut-off is used, that some countries will move between a classification of 'low', 'medium' or 'high' development. What is more significant, however, is

that the maxima and minima set in the HDI can actually alter the effective weight given to the different dimensions, and that this can result in a re-ordering of HDI-based country rankings (Panigrahi and Sivramkrishna 2002).

Returning to the work of Alkire, after identifying a core list of capabilities important to the participants in the Oxfam projects, the question of how these capabilities could be compared across activities was addressed. In order to achieve this, the impact of the activity on a particular capability was ranked relative to what was possible in the situation. The ranking was given from zero to five, with zero being no change and five being intense change relative to the context. A negative score was given if things had got worse. To enable comparisons of overall functioning level, a second kind of qualitative scoring was developed. In this ranking, focal capabilities that activities were expected to have affected were identified, defined, and a scoring system was constructed that reflected the spectrum of possible levels of achievement for each functioning. The scores an activity had on this scale enabled comparison between the levels of achieved functioning in different communities. For example, the criteria for ‘women’s participation’ are reproduced in Table 6.7 below (Alkire 2002, p230).

Table 6.7: Criteria for Scoring Women’s Participation

Functioning	Criteria for overall ranking
Women’s actual participation in the community	0 no awareness 1 awareness of the activity and some surveys/ consultation 2 some meetings/ interaction with activity initiators; some ‘contribution’, responsibility by a few people 3 some responsibilities in more than one stage/ regular meetings 4 significant involvement in different stages/ significant responsibilities and ownership/ big ‘contribution’ 5 community working in more than one sector; taking initiatives; able to mobilise resources

Having identified and defined the five functionings as part of the ICEPOP study, the next step for Grewal *et al.* was to define levels of achievement for each of the five functionings; for example levels of achievement for the functioning ‘control’ were defined as:

- I am able to be completely independent
- I am able to be independent in many things
- I am able to be independent in a few things
- I am unable to be at all independent

The next stage in the ICEPOP study was to include the five functionings and their respective levels into an instrument and develop a scoring system. In order to obtain values for their functionings Coast, Flynn *et al.* (2008) used discrete choice experimentation in combination with best-worst scaling. A pilot of the discrete choice experiment has been completed. Respondents were presented with 16 scenarios; for each scenario, respondents were offered a set of attributes at different levels, and were asked to choose which attribute they considered best and which was worst within that scenario. In interviews, a total of 255 full and useful responses were obtained for the valuation exercise; mean age of respondents was 74.6 years.

Coast, Flynn *et al.* (2008) made the decision to anchor their scale/index such that zero is equivalent to having no capability; although someone who is dead will clearly fit into this category and so will be at zero, zero is not strictly interpreted as dead. One is the best level of capability possible, as measured by the ICECAP instrument. “Having all of the love and friendship that I can have” is the highest rated attribute level; “Being unable to be at all independent” is the lowest rated attribute level (and has a negative sign).

Klasen (Klasen 1997; 2000) has used the capability approach to analyse poverty and inequality in South Africa. A novel method is used by Klasen to assess the achievement of capabilities. Klasen selects 11 capabilities and devises a scoring system to indicate the level of achievement of each. These capabilities and the scoring is summarised in tabular form and a section of the table from Klasen’s article is reproduced below (Table 6.8) for six of his 11 capabilities.

Table 6.8 Components of a Composite Measure of Deprivation (Klasen 2000, p40)

Component	Description of indicator used	Score (1 signifying most deprived, 5 least)				
		1	2	3	4	5
Education	Average years of schooling of all adult household members	<2	3-5	6-9	10-11	12+
Income	Expenditure quintiles	Poorest quintile	Quintile 2	Quintile 3	Quintile 4	Richest Quintile
Wealth	Number of household durables (e.g. vehicles, phone, TV)	0-1	2-4	5-7	8-10	11+
Housing	Housing characteristic	Shack	Traditional dwelling, hostel, outbuilding	Combination of buildings	Flat, maisonette	House
Water	Type of water access	River, stream, dam, standing water	Rainwater, protected spring, well, borehole	Public standpipe, water tanker	Piped water on premises	Piped water inside house
Sanitation	Type of sanitation facilities	No toilet	Bucket	Latrine	Chemical toilet, flush toilet outside	Flush toilet inside

As can be seen from Table 6.8, each indicator is scored on a scale of one to five. An effort has been made by the author to ensure that a score of five represents the best possible condition. A score of one is an indication of severe deprivation. By assigning scores, the differences in levels of achievements are interpreted cardinally (Klasen 2000). The scoring is a simple way of presenting different qualitative observations on a standard index.

Two procedures were used to derive a weighting of the various components of the index. One derived weights from the data itself based on principal component analysis. The other method was to calculate the total deprivation index as simply the average score of all individual components. The author states that the two procedures yielded virtually identical results.

Anand and van Hees (2006) compared achievement in terms of personal objectives, health, intellectual stimulation, social relations, pleasant environments and social projects, with respondents' self-reported satisfaction with their own capabilities. Each question used in the questionnaire had seven levels. A response of one represented a high degree of satisfaction with a capability and a response of seven represented the lowest level of satisfaction; it was found that people were most satisfied with their capabilities in the areas of health and the environment. People were least satisfied in the areas of social relations and personal projects. One explanation that the authors offer for these findings is that health and environment are, to a greater extent, determined by factors outside the person's control and therefore more amenable to adaptation (Ibid. p274). One adapts because it is irrational to worry about things one cannot control. Social relations and personal projects are, however, more reflective of an individual's own choices. A notable finding was also that higher income levels are associated with lower capability satisfactions, implying a trade-off between objective improvement and subjective dissatisfaction (Ibid. p276).

Data from the third version of the questionnaire refined by Lorgelly *et al.* was analysed in phase three of the study and in this phase an attempt was made

to generate an index of capability. The authors chose to assign equal weights to each the ten capabilities, which was felt to be the simplest approach to take (Lorgelly, Lorimer et al. 2008).

6.5 Summary

It was suggested in the introduction to this chapter that the best way in which to begin to operationalise the Capability Approach was to follow a series of five steps. These five steps were listed as:

1. A decision as to which of those concepts of advantage listed in chapter five (Figure 5.2) to focus on.
2. A decision as to whether to focus on functionings or capabilities.
3. Identifying the objects of value.
4. Judging achievement in terms of a particular functioning/ capability.
5. Establishing the value of these objects.

The literature was searched to find examples of studies in which the capability approach has been applied in practice. It was found that although there is an abundance of examples of valuable functionings and capabilities provided in the literature, ranging from educational attainment to being able to live with concern for the world of nature, there are few, but growing numbers of cases, in which a context specific list of capabilities has been developed through participatory methods. Where functionings and capabilities have been obtained through participatory methods it has been noticed that the boundaries between functionings and utility appear to be less clear-cut than is apparent from the theoretical literature on capability. It was concluded, however, that the capability approach can handle complex notions of utility by simply treating different mental states as examples of functionings.

There are also only a small number of cases where an attempt has been made to weight capabilities and to combine capabilities into a single index/number; where this has been done capabilities have most often simply been given equal weight. Only one group (Coast, Flynn et al. 2008) have

attempted to obtain values for capabilities which reflect preference over them. Although in this case, focus only switched to capability after an assessment of the initial findings.

Although many of the cases discussed in this chapter have contained functionings relating directly to health (such as infant mortality, life expectancy, and physical health), assessment of the achievement of these functionings was most commonly based on national statistics, in which they indicate the level of economic development or general well-being across entire nations. Only three of the cases discussed in this chapter have come from the discipline of health economics (Work by Grewal, Coast *et al.*, Anand, Hunter *et al.*, and Lorimer, Lorgelly *et al.*). It was argued in chapters two and three that it is quality of life rather than health in isolation that is important to patients with chronic pain. Work by Grewal *et al.* (2006) further indicates that it is the loss of ability to pursue quality of life attributes that is important to individuals, which is exactly what the capability approach is designed to capture.

Where the capability approach is being used within the discipline of health economics it is being applied in the context of public health and to assess quality of life in older people, which makes its use with chronic pain sufferers the first in the context of chronic illness.

As stated in section 6.2, several papers were added in to Table 6.4 which became available after the period of the search (which was conducted during March to May 2007). Papers were identified by monitoring journals such as Health Economics and the Journal of Human Development and Capability, and some conference papers and publications were sent directly to the author via contacts working in this area. The author is also aware of two recently started and ongoing projects applying the Capability Approach within health. First Greco *et al.* (2009) are using the Capability Approach to assess women's well-being in Malawi; the conference paper cited discusses proposals for the study design. Second, a research team from the Universities of Aberdeen and East Anglia are currently using the ICECAP-O measure (developed by

Grewal, Coast *et al.* (2006)), along with other measures, in a feasibility study involving chronic pain patients. As these two studies are ongoing, results are not yet published.

CHAPTER 7: Exploring Possible Methods for Weighting Capabilities

Several lists of capabilities have been developed within health economics. In addition to the capability-questionnaire developed for this thesis there are those of Coast *et al.* and Anand *et al.* Yet it has been noted that the process by which values should be elicited for capabilities is currently unresolved (Coast, Flynn *et al.* 2008). Indeed it was found in chapter six that it is relatively unusual to derive weights for capabilities. In eight of the 15 studies reviewed no attempt was made to determine how valuable individual capabilities/functionings were relative to each other. In six studies capabilities/functionings were simply each given equal weight. Given the fact that it is so uncommon to value functioning/capabilities it is unsurprising that there is no accepted method of doing so beyond simply assigning equal weights (which is what Anand, Lorgelly *et al.* have opted for at present).

This chapter is therefore concerned with addressing the issue of how to determine how valuable the selected capabilities and how they should be weighted.

It has been suggested in chapter five that simply assigning equal weights to capabilities is not appropriate in the context of health. The motivation for using equal weights in the capability literature is different from what we are trying to achieve. It has been suggested by Nussbaum that certain capabilities are so important that there should be no scope for tradeoffs between them; there is no way of compensating for the deterioration of one capability by improving another (Nussbaum 2003). Within this project, however, there *is* a desire to incorporate and accurately reflect the tradeoffs people are willing to make between capabilities, for two reasons. First, if the capability states defined by the instrument are not valued according to the preferences of the people who are or could be affected by chronic pain then it leaves the exercise open to claims of paternalism. Second, the reality is that health care and other public sector budgets are limited and those interventions which are funded should result in the greatest improvement in the most important capabilities. It has to be acknowledged therefore that

some capabilities may contribute more to a good quality of life than others. In only one study (Coast, Flynn et al. 2008) was a method adopted to determine preference weights; in this case, discrete choice experimentation was used.

What *has* been measured is performance on individual functionings, which tends to be measured in natural units, such as life expectancy or educational attainment. An alternative approach, adopted when things are considered for which there is no natural unit of measurement has been to assign arbitrary scores to levels within functionings, such as the scale of one to five used by Klasen (2000). In the work by Klasen, a score of five represented the best standard or condition, and a score of one indicated severe deprivation for dimensions such as 'type of water access' for households in South Africa. Alkire (2002) uses a kind of qualitative scoring system similar to that used by Klasen, to allow comparisons of the level of achieved functioning associated with different projects in different communities. In short, a scoring system was constructed that reflects the spectrum of possible levels of achievement for each functioning, zero for the worst, one for the next best level etc. It would seem unlikely though that the levels within the functionings on the questionnaire will be equidistant in terms of severity and so ideally the levels within the functionings should be valued on a scale, with values reflecting some form of public consensus.

With no common or gold standard technique apparent from the literature, some preference-based valuation method will need to be chosen to weight the capabilities for this study. It may be useful, therefore, to set out a list of minimum requirements of a valuation method along with any desirable properties.

7.1 Requirements of a Valuation Method

When it comes to valuing states for use in economic evaluation full comparability is required, and for full comparability a valuation method must produce a set of values which lie on an interval or ratio scale (Dolan 2001).

An interval scale provides information on how far apart those states are in terms of severity but does not indicate the absolute magnitude of severity. A ratio scale is achieved when the distance from zero is known for at least one state, and thus absolute severity can be determined for all states (Ibid). The three most widely used methods that generate valuations that lie on an interval scale are the visual analogue scale (VAS), the standard gamble (SG) and the time trade-off (TTO).

It is often noted that because no probabilities are involved in the measurement task, the VAS and TTO techniques produce value scores, not utilities (as discussed in chapter three). Torrance *et al.* (1996) explain that:

Values are cardinal preferences for use in situations that do not involve risk (uncertainty) and therefore are measured with questions that do not involve risk. Utilities are recommended for applications that do involve risk, and thus are measured with questions that do incorporate risk...

It is commonly believed that utilities are most appropriate, as almost all decisions about healthcare are made under conditions of uncertainty. However, it is argued by Dolan that “the appropriateness of a valuation method is determined by its ability to act as a proxy for utility and not by its capacity to model the situation being valued” (Dolan 2001). Here the valuation method is simply required to act as a proxy for well-being (or happiness and desire) and not elicit Von-Neumann Morgenstern utility. It may be unreasonable to expect to model uncertainty realistically in framing a valuation question, especially without overburdening the respondent with technical detail. If we ask a patient to complete a questionnaire which will, in turn, define their capability state, then we have an actual snapshot of their well-being. What is perhaps more important is that we can accurately value this certain (albeit for respondents on the valuation exercise, hypothetical) Capability state and scale this against all other possible states. In short, it would appear unnecessary for the purposes of this work to state as a requirement that the valuation technique must generate Von-Neumann Morgenstern utilities. What is more important is the reliability, simplicity, and the practicality of valuing Capability states.

The challenge faced is not simply one of how to weight capabilities. Instead the potentially more complicated question is how to value capability states, for weighting only the capabilities would not in itself present us with a means of 'scoring' responses to the questionnaire. We must either find a way of valuing the end-points, i.e. each of the many capability states, or in addition to weighting the capabilities we must value the levels on each and specify a means of aggregating across them.

It is also worth noting that whereas the SG and TTO techniques are based around choice, VAS is not. While in health economics, choice based techniques, such as the SG and TTO, are considered to be theoretically superior, we have seen in previous chapters that Sen has been critical of the interpretation of utility in terms of choice (Sen 1987) and so it may be desirable, if possible, to explore a choice-less technique.

Recall from chapter three that the case can be made either in favour of eliciting values from patient groups or alternatively from non-patients. The issue with eliciting values from patients is that they often adapt to their ill health over time. The value assigned to a health state is, however, likely to be an accurate reflection of how good or bad it is to *be* in that that health state. Non-patients, on the other hand, tend to focus on the shock of being newly debilitated and the negative impact of being in that health state. It was suggested in chapter three that this problem will be more exaggerated the less informed the non-patient respondents are. When presented with capability states respondents should have fuller descriptions of what their quality of life would be like in those states, expressed in terms of their ability to conduct everyday tasks and roles. In other words, respondents are asked to value scenarios which are clearly described in terms of activities, roles and abilities that they are familiar with and which already form part of their everyday lives. Respondents are not given scenarios described in terms of health and left to imagine what effect ill health would have on their daily life.

The case made in chapter three for eliciting values from non-patients was that certain patients have been found to tolerate ill health and not seek treatment for easily treatable conditions (Sen 2002). Sen suggests that this may more frequently be the case for those who have low levels of education and have spent long periods of time trapped in poverty and unfavourable social conditions. A non-patient will be just as susceptible to adapt to their own personal conditions and their valuations may well reflect this. It would seem more objective, however, to use a societal value (the average of many people's valuations) rather than one individual's assessment of their own well-being. It is perhaps not so much a requirement of the valuation technique itself, but it is felt therefore that the valuation exercise should be completed by non-patients.

Having established this general list of requirements and desirable properties, possible contenders for the final valuation method will be discussed, in turn, in the following sections. First, however, a number of possible choices have been introduced and it may be helpful to formalise these below. Even from this short introduction, it is apparent that at least three decisions need to be made:

1. To elicit values from patients or from a sample of the general public, or alternatively to give the capabilities and the levels within questions equal weight.
2. To use a choice-based technique (TTO, SG) or a choice-less approach (VAS, possibly Willingness-to-pay).
3. The third choice is a choice over *what* to value:
 - a. A selection of the many end states (the capability states) defined by the questionnaire and then use some form of model to derive values for the remaining states based on this data. I.e. a process similar to that used in order to arrive at the EQ-5D tariff values.
 - b. The capabilities and/or the levels within questions and then formulate a model to derive values for capability states based on this.
 - c. Some method falling on the spectrum between these options.

It has already been decided (see discussion in chapter three) that values will be elicited from a sample of the general population. It has been suggested that a choice-less technique may be most appropriate in the context of capability. Sen has been critical of interpreting utility (well-being) as choice and the assumption that respondents will always choose the option that they prefer.

7.2 Willingness to Pay

One option would be to use willingness to pay, the advantage of which is that benefits are expressed in the same units as costs. There are, however, at least four problems with willingness to pay, each of which was discussed in chapter three. First, part-whole bias can affect the results of WTP studies. Second, some respondents object to valuing health in monetary units and WTP studies therefore tend to include protest votes, where respondents report being unwilling to place a monetary value on health. What respondents may find more uncomfortable than placing a monetary value on health is being asked to place a monetary value on something like friendship; they may also find it odd to be asked, for example, to state how much they would be willing to pay to return to work. What is more, if respondents are asked to place a value on returning to work they may simply give the wage rate they expect they would receive. The lost wage is often included already as an indirect cost and the wage rate may not fully represent the personal gain to the individual in terms of feeling useful, being with colleagues and experiencing greater self-respect.

Third, willingness to pay is dependent on ability to pay and this creates issues in terms of equity and in terms of how good money actually is as a proxy for utility. This is important in the context of this study because the most obvious approach would be to ask respondents how much they would be willing to pay for a hypothetical treatment which would transform them from some dysfunctional state on the questionnaire to some state of improved quality of life. Some states on the questionnaire are highly undesirable (such as being

completely dependent on others to dress, shower etc, no longer feeling included as a member of any social group, and never being free of pain) and it is quite feasible that as people may be willing to pay large sums of money to escape these situations, their budget constraint will begin to have an impact on their responses leading to insensitivity of the measure (Smith 2005).

We could frame the WTP question in order to generate an option value; here members of the public (potential taxpayers) would decide upon the maximum premium they would be willing to pay to be guaranteed access to an intervention which, should they find themselves in a worse Capability state at some future point, will restore them to their current – or at least to some improved – state. One would expect budget constraints to have less of an impact in this case.

Finally, it is questionable whether the theoretical basis of cost-benefit analysis fits at all comfortably with the Capability Approach. WTP is a cardinal measure which is used to represent a change in utility. It has been noted in previous chapters that it is not immediately clear how to define utility, and that the range of concepts which are labelled as utility are generally unduly narrow and highly subjective.

7.3 Discrete Choice Experiments

A second option would clearly be to adopt the same methods as have been used by Coast, Flynn et al (2006), i.e. Discrete Choice experiments with best-worst scaling. The authors suggest that DCE may be appropriate in the context of capability because values and not preferences are elicited:

Although best-worst scaling can be viewed as a choice experiment, respondents are asked only to specify which attribute levels they think are the best and the worst. Hence, it can be argued that values and not preferences are elicited, because individuals are not asked to trade one thing for another. Thus, the best-worst scaling approach may come closer to eliciting values that would satisfy Sen's interpretation than measures produced by other methods... (Coast, Flynn et al. 2008)

Coast, Flynn *et al.*, state that a major strength of discrete choice experimentation (DCE) is that it allows the estimation of the importance of the entire attributes/capabilities as well as the levels of those attributes. Values used in the ICEPOP study are scaled such that Capability State 11111, which can be interpreted as the absence of capability, is equal to zero and full capability, the state 44444, is equal to one. It is also suggested that methods requiring only ordinal rather than cardinal responses are cognitively easier for respondents, i.e. “stating that A is preferred to B is cognitively easier than stating how much A is preferred to B” (Coast, Flynn, et al. 2008a, p34). And, ordinal tasks can still be used to make meaningful inference about underlying numerical values (Ibid.).

DCE/BWS represents one option for valuing capability and is already being explored. As work in this area has been so minimal there is a reasonable argument for continuing to search for new methods which are currently untested in the field of capability.

7.4 Multi-Attribute Utility Methods

Estimation of a tariff/scoring system for multi-attribute health status classification systems, such as EQ-5D, QWB and the HUI Marks I, II & III, has been achieved through the use of multi-attribute utility type methods. Multi-attribute utility theory is “a technique for determining a mathematical formula that will allow the estimation of preference scores for a large number of health states defined in a multi-attribute framework, based on the measurement of preference scores for a small... subset of those states” (Torrance, Furlong et al. 1995). The Health Utilities Index Marks I, II and III were developed explicitly from Multi-Attribute Utility Theory (MAUT). For the Health Utility Index Mark I, the measurement instruments used were VAS and TTO; for Marks II and III, the measurement instruments were VAS and SG.

Observed values for 42 health states on the EQ-5D were obtained by the Measurement and Valuation of Health Group (1995) using both a VAS and the TTO method. A unique model was developed by the MVH Group (“Dolan-

N3") which predicts the value of a health state from its components by attaching a value to each single deviation from good health. The model contains two other terms, one of which is active whenever any dimension of health is 'severe', and the other is a constant term which can be interpreted as the loss of value involved with being in any kind of dysfunctional state whatsoever.

It should be noted that in the case of the HUI, VAS values were converted to utilities using a power curve estimate of a risk aversion function. In contrast, the MVH Group presents 'raw' VAS and TTO scores in their 1995 report.

One option available to us for use in this study, therefore, is to adopt an MAU method using SG or TTO to value the quality of life states defined by our capability questionnaire, although as discussed in the previous chapter, these choice based methods do not fit well with the theoretical foundations of the Capability Approach.

Peacock, Richardson *et al.* (2007) have used a multi-attribute utility approach which has the potential to be adapted in order to obtain values for capabilities. Their approach was developed in the context of setting health service priorities using programme budgeting and marginal analysis, but also seems attractive for the purposes of this thesis. It should be noted that rather than valuing a subset of endpoints, as was done by the EurQol Group to derive the tariff values, the method by Peacock and Richardson involves valuing the levels within attributes and deriving values for endpoints from this. There are six methodological steps, outlined below.

Methodological Steps in the MAU approach (Peacock, Richardson et al. 2007)

1. Identifying attributes in the MAU function
2. Describing attributes
3. Scaling attribute levels
4. Quantifying Trade-offs between attributes
5. Evaluating Programmes
6. Combining attribute scores

Peacock *et al.* suggest that for stage one, an advisory panel defines attributes, which are relevant to the organisational context and the objectives of the decision makers. For stage two, the advisory panel constructs measurement scales for each attribute, against which programmes are to be evaluated. Stage three involves the panel scaling the different levels within the attributes, with their relative importance being determined on a 0-100 scale. It is suggested that 100 would represent 'excellent health' and 0 would represent 'dead'. The panel would then assess the relative importance of each attribute, through the use of a 'swing weights' method (described in more detail later). Next, the different programmes would be evaluated according to how well they perform with respect to each attribute. Finally, panel scores for each attribute would be combined using the MAU model to calculate the combined benefit score of a programme (Peacock et al 2007). A further stage is also suggested, which would involve the panel validating the results through discussion and deliberation.

This method is seemingly straight forward, and is a choice-less technique for deriving values for a set of independent dimensions. It is also of interest because of the fact that it is relatively new and, as yet, largely untested. For all of these reasons this method by Peacock and Richardson will be explored as a means of valuing our capabilities and the Capability States defined by our instrument.

7.5 Adopting the MAU approach in the context of capability

The first two steps described above are simply the stages which will be undertaken within the first phase of the study to develop the questionnaire and which will be achieved through qualitative work with chronic pain patients (see chapter eight). This is in keeping with recommendations within the capability literature to develop lists of capabilities through participatory methods.

It is envisaged that for step three the research team will ask groups of the general population to assign the top level on a particular capability a score of

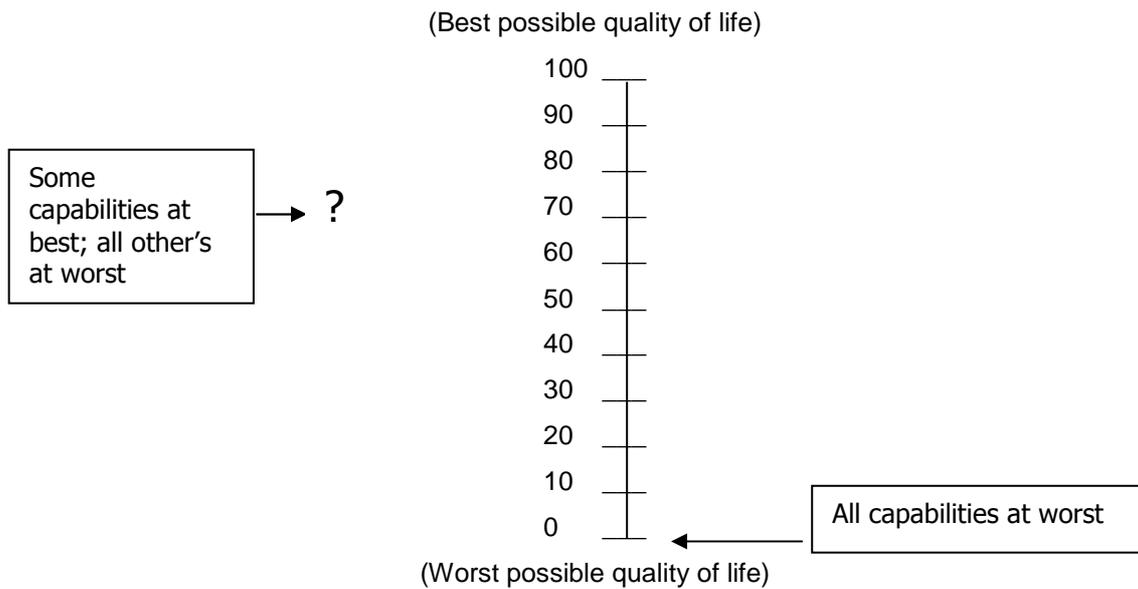
100, the bottom level a score of 0 and then place the remaining two levels on the scale somewhere between these two extremes (see chapter 10).

Each of the groups will then be asked to consider and state the relative weights of the capabilities using the swing weight method. This is a simple method in which respondents are asked how much a dimension contributes to overall well-being relative to other dimensions by comparing hypothetical states of well-being that 'swing' between the best and worst levels in each dimension. They then estimate the change in well-being that would result from changing each dimension from its worst to best level using a rating scale which has endpoints of 100 (all dimensions at their best level) and 0 (all dimensions at their worst level). Two forms of the swing weight method can be used, a 'bottom up method' and a 'top down method'. Here the bottom up method will be introduced; 'bottom up' indicating that respondents are asked to consider the relative 'goodness' of a state in which at least one capability is at its best level relative to a starting point of all capabilities being at their worst level.

Each group of respondents will be asked to consider the specific case of a person for whom all capabilities are at their worst possible extreme. This state would be represented by a 0 (all capabilities are at their worst level). They will then be asked to consider a new state in which there has been some improvement and a sub-set of capabilities have swung to being at best. The group will be asked to decide where to place this new state on a scale such as that illustrated in Figure 7.1.

The main benefit of this approach would be that the scaling of levels within capabilities and the setting of weights for the capabilities could be done following a process of public consideration and discussion, which is in keeping with the principles of the Capability Approach. What is more, the process itself is relatively straightforward.

Figure 7.1: Weighting Attributes – Bottom-Up Method



As value will be the underlying metric here and not utility, it would be more accurate to refer to the overall method as a multi-attribute value method; this reflects the fact that the values are obtained under conditions of certainty, the values still reflect personal preference.

7.6 Choosing a Functional Form for the MAV Model

The type of functions typically selected for Multi-Attribute Utility/Value models are additive, multiplicative, or multi-linear. The additive model is just the sum of the importance weighted dimensions; importance weights must be relatively small and must sum to unity (Peacock, Richardson et al. 2007). An example of an additive model would be the following utility function:

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z)$$

Where k_Y and k_Z are scaling constants. The separate contributions of the two dimensions are added together to obtain the total utility (Keeney and Raiffa 1993). The additive function implies that Y and Z are mutually utility independent; in this case, preferences for varying amounts of Y can be assessed after fixing Z at any convenient level. In other words, additive

independence means that the strength of preference in a single dimension is unaffected by other, constant, dimensions.

When Y and Z are mutually utility independent, then $u(y,z)$ can be expressed by the multi-linear representation (Keeney and Raiffa 1993):

$$u(y,z) = k_Y u_Y(y) + k_Z u_Z(z) + k_{YZ} u_Y(y) u_Z(z)$$

Where, $u(y,z)$ is normalised by $u(y_0, z_0) = 0$ and $u(y_1, z_1) = 1$.

The above multi-linear form has a strategically equivalent multiplicative representation (provided that $k \neq 0$):

$$u'(y,z) = u'(y, z_0) u'(y_0, z)$$

If it is the case that neither Y nor Z is utility independent of the other, then since mutual utility independence is a necessary condition for additive independence, none of the functional forms presented so far will be strictly appropriate.

MAU theory suggests that when two or more dimensions can independently have a large impact on overall quality of life, a multiplicative or multi-linear model should be used (Peacock, Richardson et al. 2007, p900). These models would allow extreme pain, for example, to independently reduce quality of life to zero.

Keeney and Raiffa (1993) advise us that the additive and multiplicative functions appear to be the more practical for models involving four or more dimensions. "Even when the requisite assumptions do not precisely hold over the domains of all the [dimensions], it may be a good approximation to assume they do, or it may be reasonable to integrate different additive and multiplicative utility functions over separate regions of these [dimensions]." (Ibid. p298). With individual health, community health and equity as Peacock *et al.*'s three dimensions, they use a model which is additive with respect to

dimensions other than individual health, but which incorporates a multiplicative relationship between individual health and these other dimensions. The model is calibrated so that the importance weights sum to unity. This model is referred to by Peacock et al. as the 'multiplicative weights for health model'.

$$U = U_H [1 + W_1 D_1 + \dots + W_n D_n]$$

Where U is combined utility, U_H is the utility for individual health, D_1, \dots, D_n are scores for the other n attributes on their respective measurement scales, and W_1, \dots, W_n are importance weights for those dimensions (Peacock, Richardson et al. 2007).

The Capabilities identified during the qualitative work may or may not be independent of each other. If there is independence then a simple additive model will be adequate and appropriate. Indeed, an additive model may anyway be a good approximation. It may be necessary to explore some form of multiplicative element once the data has been analysed (see chapter 10).

7.7 Summary

In this chapter various methods have been discussed which may be suitable for the purpose of weighting capabilities, although it was noted that in order to develop a scoring system for the questionnaire more must be done than simply obtaining relative weights for the capabilities. One approach would be to obtain values for the end-points (the Capability states) and to attempt to infer weights for the capabilities from this. For this task, methods such as DCE, SG, VAS, or TTO would typically be adopted. Another option would be to obtain relative weights for the capabilities and to combine these with values for the levels within them.

It was felt that there would be merit in exploring a new and choice-less approach and therefore, the second option was favoured. The chosen method will be heavily based on a method used by Peacock *et al*, which was

developed in the context of assisting the setting of health service priorities, but which – it is suggested – will offer a useful and acceptable means of deriving a scoring system for our capability-based questionnaire.

CHAPTER 8: Aims and Objectives of the Empirical Work

8.1 Motivation

Reviews of the literature on chronic pain conducted by Breen (2002) and Richardson (2005) reveal a number of important themes which are important to chronic pain patients. Ten themes were identified and listed in chapter two of this thesis, which were:

- Pain itself (intensity, frequency and location)
- legitimisation and identity
- independence
- activities of daily living
- employment
- sleep disturbance
- relationships
- social issues and loneliness
- psychological and emotional distress.

Four pain-specific outcome measures and three generic measures of health-related quality of life (EQ-5D, Nottingham Health Profile and SF-6D) were introduced in chapter two and it was noted that these instruments incorporate few of the themes listed above. Chapter four suggests that when selecting outcome measures for use in economic evaluation, researchers have so far shown more concern with using validated instruments for which 'ready' utility values exist than with the actual descriptive system. It is therefore the case that economic evaluations have incorporated little of what is important to the patient (little of which extends beyond a narrow concept of health).

As highlighted in chapter three, there are a number of other problems associated with using a narrow concept of health as the outcome in economic evaluation. First, the maximisation of health makes comparing the effectiveness of health interventions with the support provided by other

agencies difficult. Second, the maximisation of health makes the separation of equity and efficiency impossible.

As there are also a number of conceptual and methodological issues associated with willingness-to-pay, a case can be made for exploring a new framework within which to assess outcomes for inclusion in economic evaluation. This framework should offer a broad evaluative space so that as many of the impacts of ill health are included as possible and so that it is feasible for the effectiveness of health interventions to be compared alongside interventions from other agencies or government departments. Finally, the framework should be able to incorporate concerns for equity as well as efficiency.

8.2 Aims & Objectives

The Capability Approach has been introduced as a potentially useful framework within which wider quality of life could be assessed. Sen talks of equality in terms of basic capabilities (Nussbaum and Sen 1993). Essentially, society should ensure that all its citizens have the freedom to achieve various 'basic' capabilities.

Although the theory behind the Capability Approach has been described as being radically underspecified (Robeyns 2005), a number of steps were nevertheless identified as being important when operationalising the Approach. Five steps were outlined in chapters five and six:

1. A decision as to which of the concepts of advantage to focus on: (i) well-being achievement; (ii) well-being-freedom; (iii) agency achievement; or (iv) agency freedom.
2. A decision as to whether to focus on functionings or capabilities.
3. Identifying the objects of value.
4. Judging achievement in terms of a particular functioning/ capability.
5. Establishing the value of these objects.

Recall that achievement relates to observed functionings, freedom to capability. Well-being relates only to the individual's personal state, whereas agency is not restricted to the advancement of the individual's own well-being. So for example, a person's agency goals could include the happiness and well-being of family and friends, or the protection of local wildlife. To some extent the answer to the first question will dictate the answer to the second, so that if the objective is an assessment of well-being achievement then the focus is likely to be on functionings.

It was found in chapter six that it is unusual for studies to implement and explicitly discuss all of the five steps, even in the disciplines in which the Capability Approach has been developed. It can therefore be said that it is rare for the Capability Approach to be fully operationalised.

The aim of the following empirical work is to explore the feasibility of using the Capability Approach as a framework in which to assess quality of life in patients with chronic pain. The objectives are based around the implementation of each of the five steps outlined above. An attempt will be made to assess the freedom of patients to achieve important outcomes; that is to assess capability. If agency goals are found to be important to patients then they will be included in the assessment. Given that previous studies involving chronic pain patients have reported fears of being a burden or bother, it is likely that agency goals will be important. The concept of advantage being assessed is therefore likely to be agency freedom. A list of capabilities will be drawn up which are specific to the context of chronic pain, through participatory methods involving chronic pain patients. In order to judge achievement in patients in terms of these capabilities, a questionnaire will be developed for self-completion. Finally, the value of the capabilities found to be of importance will be determined using the Multi-Attribute Value Method introduced in chapter seven.

8.2 Identifying Capabilities of value to Chronic Pain Patients

It was noted in chapter five that there is debate in the development literature regarding the issue of lists of relevant capabilities. It was argued that because Nussbaum's 'central list' of capabilities was not designed for the purpose of assessing individual quality of life, and because chronic pain is unrelated to the focus of any of the studies reviewed in chapter six, that a new list of capabilities should be developed specifically for this study. Sen (2004) as well as others (Alkire 2002; Robeyns 2005b) have also argued against a reliance on a central list of capabilities. For small-scale projects, such as this, it is suggested that capabilities are selected through participatory methods (Alkire 2002; Robeyns 2005b).

Four studies were discussed in chapter six which involved the identification of capabilities through a participatory approach. Following in the spirit of these four studies, a list of context specific capabilities will be developed through the use of qualitative work involving a sample of the target population, in this case chronic pain patients. The themes already identified as being important to chronic pain patients in chapter two will be used to inform the topic guide for the qualitative work outlined in chapter eight. It should be realised, however, that differences in the way health systems are funded and structured, cultural and geographical differences and the fact that respondents will be prompted to think in terms of capability *may* mean the final list of capabilities derived in chapter nine differs from the themes identified in chapter two.

8.3 Assessing Achievement in terms of Capability

The capabilities identified in chapter nine will be used to formulate a questionnaire to assess outcome (in terms of changes in quality of life) associated with health treatment for incorporation within economic evaluation. The format of the questionnaire will depend on the nature of the capabilities identified from the qualitative work, but the intention will be to assess a respondent's achievement in terms of each capability.

It will be important to pilot the questionnaire in order to assess whether patients understand and respond appropriately to the questions and to gain an idea of the usefulness of the results. It is still unclear whether it is feasible to assess capability rather than simply functionings.

8.4 Determining the Value of Capabilities

This is a step only implemented in one of the studies discussed in chapter six (Coast, Flynn et al. 2008). The Multi-Attribute Value Method introduced in chapter seven (Peacock, Richardson et al. 2007) will be piloted as a possible method for weighting capabilities. Following discussion in chapter three of issues relating to adaptation, and that all tax payers are potential users of the healthcare system, the MAV method will be piloted with a sample of the general public.

8.5 Summary

This chapter has set out the motivation for conducting the empirical work outlined in the following three chapters. The aim of the empirical work is to explore the feasibility of using the Capability Approach as a framework within which to assess quality of life in patients with chronic pain. Results will indicate the feasibility of incorporating a score reflecting quality of life as the outcome measure within economic evaluation.

CHAPTER 9: Study Phase I: Identifying Capabilities and Developing a Capability Instrument

In this initial phase of the study, a series of focus groups were held from which information was gathered on how chronic pain impacts on quality of life. Framework analysis was used to identify a list of key capabilities from the raw data collected during the focus groups; all of these key capabilities are, to a significant extent, restricted by chronic pain. The list of capabilities was then presented to and reviewed with participants from the focus groups during a series of individual interviews. The final list of capabilities was then developed into a questionnaire. The process of using qualitative work to arrive at a list of higher-order capabilities is outlined in this chapter.

9.1 Selecting a Means of Collecting Qualitative Data

In the four studies identified in chapter six, in which lists of capabilities were obtained through participatory approaches, either individual interviews and/or focus groups were used to arrive at a list of valuable and context specific capabilities. A decision over whether it was most appropriate to use individual interviews or focus groups for this study was not immediately obvious. A major advantage of using focus groups is that this method provides a large amount of information in a short period of time, and at relatively low cost. A disadvantage is that it cannot be assumed that individuals in a focus group are expressing their own definitive individual view as they are speaking in a particular context (Gibbs 1997). Focus groups can also be intimidating at times, especially for shy or inarticulate members, and there is a risk that the discussion will be dominated by one individual. Kitzinger (1994), however, states that “it should not be assumed that groups, by definition, are inhibiting relative to the supposed ‘privacy’ of an interview situation. ...the less inhibited members of the group ‘break the ice’ for shyer participants or one person’s revelation... encourages others to disclose.” (p111).

What makes the discussion in a focus group more than the sum of separate individual interviews is the fact that the participants both query each other and

explain themselves to each other (Morgan 1996, p139). Such interaction provides direct evidence about similarities and differences in the respondent's opinions and experiences as opposed to reaching such conclusions from *post hoc* analysis of separate statements from each interviewee (Morgan 1997). A reliance on group interaction, however, means that focus groups require greater attention to the role of the moderator and provide less depth and detail about the opinions and experiences of any given participant.

Given a limit on time and resources, and a desire to obtain a list of general capabilities on which chronic pain has an impact, rather than detailed information about each individual's quality of life, a decision was made to conduct focus groups as opposed to individual interviews.

9.2 Focus Group Design

There are recommendations that a full focus group should consist of between six and 10 participants (Gibbs 1997; Morgan 1997; Finch and Lewis 2003), and that smaller focus groups should consist of between three and six participants (Kitzinger 1994; Morgan 1997). Smaller groups make it easier for moderators to manage the active discussions that often accompany high levels of involvement and less neutral topics (Morgan 1996, p146). Given that participants in this study were asked to discuss their own experiences we expected participants to have quite high levels of input, and therefore a decision was made to run smaller groups.

Participants for the focus groups were selected so that each group was as homogenous as was reasonably possible. Homogeneity not only allows for more free-flowing conversation among participants but also facilitates analyses that examine differences in perspective between groups (Morgan 1997, p35). With sensitive topics a shared experience is more important and sensitive topics therefore leave less scope for diversity (Finch and Lewis 2003). It is common to split participants according to gender and so all male and all female groups were held.

Participants were also allocated to a particular group according to age; specifically there was an intention to divide participants into groups according to whether they were over the UK retirement age. The reason for splitting participants according to their age was to allow comparison between the results from this study and results from the work conducted by Grewal et al. (2006). There was also the consideration that older participants were likely to have retired from work, were unlikely to have young children and were more likely to be experiencing other health problems in addition to their chronic pain. When female participants were allocated to groups in practice, they were allocated to the older group if they were aged 60 and above *and* were retired; this reflects the lower retirement age for women, and the fact that the younger groups were divided according to those who were in employment and those not in employment.

High rates of unemployment and difficulties carrying out paid work were issues discussed in chapter two. It was expected that those participants not in work were likely to be unable to work due to their pain and may report distinctly different experiences to those who were still able to work. It was also considered that participants not working due to personal choice, for example to concentrate on full-time child care responsibilities, may also report different limitations, concerns or difficulties to those who were employed.

Participants were therefore divided into six distinct segments, which were:

1. Male, aged 65 years or above and retired
2. Female, aged 60 or above and retired
3. Male, aged 18-65, in paid employment
4. Female, aged 18-60, in paid employment
5. Male, 18-65, not employed
6. Female, 18-60, not employed

Focus groups were conducted in one of three rooms at the James Paget NHS Hospital, Norfolk. Rooms used were: the Barton room (in the Education and Training Centre), the waiting area (at the Pain Management Clinic), and the Rockland Room (in the Burrage Centre). The Barton Room and Rockland

Room are seminar rooms with seating for 15 and 30 people respectively. The waiting area at the Pain Management Clinic was used in the evening so that staff and other patients were not present during the group discussion.

At the start of each group the moderator (myself) welcomed participants and ten minutes was allowed for refreshments. During these first ten minutes participants were also informed by the moderator that an audio recording would be made of the discussion and asked to treat the comments of other participants as confidential upon leaving the venue. The recording was then started and the moderator worked through a topic guide, which is included in Appendix A. The groups were first asked to list any activities, freedoms or roles which they felt were valuable and which contributed to a good quality of life. They were then asked if their chronic pain restricted their freedom or ability to do any of the activities or roles from the first list. As participants responded they were asked to provide further detail on how each of these restrictions impacted on their quality of life and why/how each of the activities or roles were of value to them. This process of moving from a set of activities/roles to the uncovering of their underlying personal value drew fairly loosely on the technique of laddering, a specific form of interviewing which takes a structured approach to the data gathering process (Miles and Rowe 2004). The Laddering Technique involves the participants being asked why each activity/role (in turn) is important, using a simple probe like “why is this important to you?” the reason given in response to this probe is further questioned, so that essentially the technique involves a sequence of why-questions (Miles and Rowe 2004; Sorensen and Askegaard 2007).

The groups were next asked if there was anything which they had not included on the first list because their pain meant that they were unable to do these things. There was then a more specific focus on the areas of freedom, independence and relationships. Each group was scheduled to last for between one to one and a half hours.

For five of the six groups the moderator was accompanied by an observer (either Richard Smith or Angela Robinson) who made brief notes during the

discussion and, where appropriate, asked participants additional follow-up questions based on their comments in the groups. On only one occasion was an observer not present and this was because only one participant from the segment was able to attend; in this case the topic guide provided the structure for an individual interview. Focus groups were held between 20th November and 29th November 2006.

9.3 Initial Recruitment

All participants were recruited through the Pain Management Clinic at the James Paget NHS Hospital. Existing patients (i.e. patients who had attended at least one previous appointment at the clinic) who were due to attend a routine appointment to see a Consultant Anaesthetist were identified and sent an invitation letter and information sheet (see Appendix A) by post one week prior to their appointment. When patients attended the clinic for their appointment they were approached and asked if they had received information about the study and if there were any questions that they would like to ask about the study. They were then asked if they would like to participate. Patients who chose to participate were asked to sign a consent form and to provide some preliminary information (see Appendix A). The preliminary information sheet simply recorded gender, age and employment status, to allow for the allocation of the patient to a particular segment and hence to a particular focus group.

Invitation letters and information sheets were approved as part of the overall design for this phase of the study by the Norfolk Research Ethics Committee. Scientific Review was conducted by the East Norfolk and Waveney Research Governance Committee. Invitation letters and information sheets were sent out from the Pain Management Clinic on the Hospital's headed paper.

All 83 patients attending the Clinic within the three week period between 23rd October 2006 and 10th November 2006 were invited to participate in the study. The inclusion criteria for this phase of the study were as follows:

- All existing patients must be attending the pain clinic due to chronic non-malignant pain.
- Patients must be able to read and understand English, in order to understand the Information Sheet and give consent.
- All patients must be over the age of 18 years.

Two patients were excluded on additional grounds. The first was excluded at the stage at which invitation letters were sent out as their address was given as a local prison and the ethical approval that we had did not cover research involving prisoners. The second patient felt, on reflection, that participating would have caused them too much distress and withdrew consent.

9.3.1 Participant Characteristics

Table 9.1 shows that, from the 83 patients who were invited to participate, a total of 33 patients (40%) indicated that they would be willing to take part in a focus group, although only 16 of those who had signed a Consent Form actually attended one of the groups. The most common reason patients gave for deciding not to participate was that they did not wish to discuss their chronic pain in front of a group.

Given that some patients did not want to discuss their pain in front of a group, interviews could have been offered as an alternative in order to maximise the diversity of participants included. However, a series of individual interviews would have involved a significant diversion from the study proposal approved by the Ethics Committee. Furthermore, in research conducted by Richardson (2005), which did involve individual interviews, respondents were still reluctant to discuss their pain and felt it necessary to defend themselves against talking too much about pain. It is therefore possible that those patients who declined to participate in the groups would also have found participating in an interview uncomfortable or even distressing.

A minority of those who signed Consent Forms could not be contacted in order to be allocated to a group, those who were successfully contacted could

not always all make the same date and time and a further number simply failed to attend on the day. Table 9.1 also gives details of the make up of the focus groups.

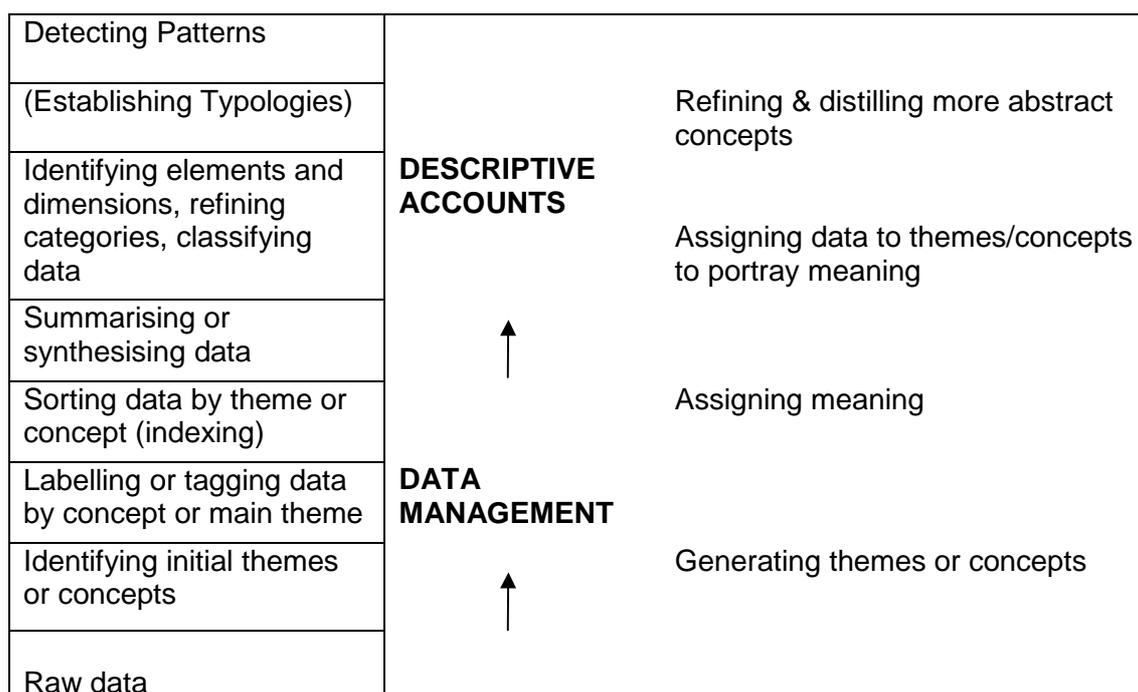
Table 9.1 Focus Group Participants

Segment	Number of Consent Forms Obtained	Number of participants	Mean Age of Participants
1 (Male, 65+)	4	2	74
2 (Female, 60+)	6	5	68
3 (Male, 18-65, Employed)	5	2	53
4 (Female, 18-60, Employed)	8	1	33
5 (Male, 18-65, Not in Employment)	5	3	45
6 (Female, 18-60, Not in Employment)	5	3	52
<i>Total</i>	33	16	

9.4 Framework Analysis

Data from the focus groups was analysed using framework analysis, which is a method of classifying and organising data according to key themes, concepts and emergent categories (Ritchie and Lewis 2003). Ritchie and Lewis set out a structure for Framework Analysis, which they refer to as the 'analytic hierarchy', reproduced in Figure 9.1 (Ibid. p212).

Figure 9.1 The Stages and Processes involved in Framework Analysis:



9.4.1 Data Management

In this study the ‘raw data’ simply took the form of audio tapes, together with brief notes made during the focus groups by either Richard Smith or Angela Robinson. At first, three of the six tapes were played in their entirety and a note was made of any themes arising. This involved selecting key words or phrases, such as ‘family’, ‘sport’, ‘frustration’, and ‘reliance on others’. An ‘index’ was then developed whereby each theme was grouped under a smaller number of broader ‘main themes’ such as ‘Employment’, ‘Mood’, ‘leisure & Sport’, etc. Main themes are listed in Figure 9.2. Themes in italics were later incorporated into other themes (see Figure 9.3).

Having constructed this initial conceptual framework, the next task was to apply it to the raw data, a process Ritchie and Lewis refer to as ‘indexing’. All six of the tapes were played through in turn and for each short section of the discussion the question asked was “which theme is being discussed?” Each time one of the themes was mentioned in the discussion the location was recorded by making a note of the reading from the tape counter. If new themes arose then these were added under an appropriate ‘main theme’.

Because the raw data was in the form of audio tapes it was not possible to physically sort the data by theme or concept.

Figure 9.2 Main Themes (Initial)

- Employment
- *Holidays*
- Family
- Mood
- Understanding (of one's own condition)
- Understanding (shown by others)
- Leisure & Sport
- Comfort
- Mobility
- Sleep
- *Socialising*
- Isolation
- Relationships
- Independence
- *Self-care*
- Medication
- *Everyday tasks*
- *Mental Well-being*
- *Physical Well-being*
- *The Future*
- 'Other'

Table 9.2 gives a brief example of the 'sub-themes' which were grouped under the main themes of 'Employment', 'Isolation' and 'Everyday Tasks'. It should be noted that themes and sub-themes were identified if they were discussed in *any* of the six groups, but each particular sub-theme is unlikely to have been discussed in *all* six groups. Within the main theme 'Isolation', 'not being able to tell people how you feel' is an example of a new sub-theme which was added into the index as it arose in the discussion within the fourth group/segment.

discussed with Richard Smith and Angela Robinson, who had been present at the focus groups and had taken their own notes.

Figure 9.3 Refined List of Main Themes

- Family
- Employment
- Leisure & Sport (incorporating 'Holidays' & 'Socialising')
- Isolation (incorporating 'Socialising')
- Relationships
- Mobility (partly incorporating 'Physical Well-being')
- Identity
- Independence & Control (incorporating 'self-care' and 'Everyday Tasks')
- Sleep
- Other People's Understanding & Acceptance
- Individual Understanding & Acceptance
- Medication (partly incorporating 'Physical well-being')
- Pain & Discomfort
- Mood ('Mental Well-being')

Examples of the summaries of the main themes 'Employment' and 'Independence & Control' are provided in Figures 9.4 and 9.5.

Figure 9.4 An initial summary of the main theme 'Employment'

Respondents reported having to change their job, limiting their hours, being unable to do certain tasks at work (either due to the physical requirement or due to a lack of concentration) and, in some cases, having to give up work altogether.

"I have had problems, erm, regarding my pain, in that I've lost a couple of jobs actually from having too much time off because of the pain" (Female, Employed).

"Can I ask? did you give up work because of your pain? (PK)

"Yeah, I had to, I, er, the tablets I'm on, they stop, they took my heavy goods licence away from me..." (Male, Retired).

Work was discussed as a source of disposable income ("having wages in your pocket"). Working meant that respondents did not have to rely on benefits, which they did not feel were adequate enough to fund a good quality of life.

Respondents felt that it is important to make a positive contribution ("doing something worthwhile"), and hence achieve self-respect. Work was seen as a way of remaining active; respondents feared being viewed as lazy or as "scroungers" if they did not work.

"Working is important just to do things that aren't mind numbing, if you know what I mean..." (Female Employed)

"What upsets me is that I'm unable to work. I'm only 39 years old, I think to myself, is that my life already. It makes me depressed." (Male, Not in Work)

Work is also an opportunity to meet people and leads to social interaction.

"My job was a very social job as well, it was very long days, but it generally ended across the road in the Coach & Horses, because that's what you do, you know, the cliché: 'work hard, play hard'. And that whole thing stops, so you haven't got any of that. So the people that you worked with were your work chums, ... and all of a sudden you haven't got anything, you don't keep up with the people generally that you worked with, I think, because you don't have the same interests and things in common because you haven't got that ongoing thing of what's happening every day" (Female, Not in Work).

Figure 9.5 An initial summary of the main theme 'Independence & Control'

Respondents reported difficulties relating to self-care (showering, dressing, etc)

"This morning, I got up - 5 o'clock - I took my first pain killers, went back to bed again so that I was ready to get up to have my shower at half past six, or else, by the time you start taking them they haven't taken effect and you're trying to move around. So, yeah, you've got to think ahead..." (Female 18-60)

Respondents disliked having to ask others for help (e.g. with cooking, shopping or housework). Respondents felt that more planning was needed to manage their lives and their home as they were restricted in their ability to do things such as walk to the local shop. They also felt that the unpredictable nature of their pain led to uncertainty and to a lack of control.

"You're relying on everyone all the while, and that's horrible, to rely on someone all the while" (Male, Not in Work).

"I can't pick up saucepans and this and that, I live on pizzas" (Male, Not in Work).

"... say you want some milk from the shop, that's only just down the road, but I can't walk and get it..." (Female, Retired)

"Now, I have to instruct someone when to take something out the freezer, erm, and, it's very frustrating, trying to explain to somebody how to do something in the kitchen..." (Female, 18-60)

"I used to be a very independent person, who went out a lot and who walked a lot; in the good weather and the bad, and now for instance I can't take the dog for a walk in the winter..." (Female, Not in Work).

Some respondents reported being unable to use public transport and/or drive due to pain and/or embarrassment; others reported being unable to go out in bad weather.

"... obviously I take a lot of pain killers, and then I can't drive" (Female)

It is suggested (Ritchie and Lewis 2003) that descriptive analysis involve three stages: categorisation, classification, and detection. Categorisation involved assigning 'labels' to the data which could move away from the original text and could bring together features which were originally described differently. There was also a classification and hence a separation of participants' feelings and opinions, for example anger or confusion, relating to the *cause* of their pain, from feelings and reported experiences relating to the *impact* of the pain on their quality of life. Greater emphasis was placed on the impact of the pain.

It was considered important to detect patterns (linkages) between themes and to challenge the location of each theme under a particular label. Some themes could have been placed under more than one label, although the labels themselves were intended to be unique/independent of each other. So, for example, struggling to undertake daily activities will clearly impact on the level of independence that an individual can enjoy, but needing the help of others may also to some extent define the individual's identity and highlights the fact that they are a person with an illness.

Detection involved looking at each theme and looking for/at the range of perceptions and experiences relating to that theme, both within groups and across groups. So, for example, a participant taking morphine might report very different experiences than a participant receiving a different type of treatment, even though the participants are in the same segment. On the other hand, participants in the male 18-65 segments may have similar hobbies and interests, but hobbies and interests which are very different to those enjoyed by females in the older group. The main purpose of detection in this particular study was to ensure that themes were not drawn out which were too specific to one respondent or one segment; it has to be remembered that the aim is to develop dimensions for a questionnaire which are general enough for all pain patients to understand and identify with.

Where there were differences between individual responses these appeared to be mostly driven by differences in gender and pain severity. For example, when it came to relationships and identity, female respondents seemed to be worried more about their pain making them feel older than they were and making them less attractive to their partners. Male respondents tended to place importance on the ability to do tasks that they felt defined the role of the man, such as carrying heavy items, cutting the grass, and washing the car.

Respondents who had suffered with very severe pain for a long time seemed to suffer more side-effects from very powerful medication and were often resigned to the fact that their pain would get worse. It was noticed that they tended to speak of depression and reported fearing the future. Respondents

whose pain had started more recently tended to express a sense of hope that a cure would be found and that their condition would improve in time.

Many of the patients had young families, either young children or grandchildren and so this was an area that was given a considerable amount of attention. There is a danger that parts of the questionnaire will appear less relevant to respondents who do not have children. The role of parent or grandparent was included because regardless of whether respondents actually do have children the role of parent is an important one and any disruption to the parent-child relationship is surely significant and undesirable.

Little emphasis was placed on establishing detailed typologies as – once again – the idea was to develop a list of capabilities broad enough to be relevant to all chronic pain patients, not to define patients according to certain sets of characteristics. Various studies discussed in Chapter Two (Blomqvist and Edberg 2002; Raheim and Haland 2006; Richardson, Ong et al. 2006) have done this, but establishing a typology was not the objective of this study.

9.4.3 Refining Categories & Classifying Data

As stated above, categorisation involved assigning 'labels' to the data which could move away from the original text and could bring together features which were originally described differently.

It was felt to be the case that 'going out to work' actually contributes to respondents achieving self-respect, facilitates interaction with other people and is a means of remaining active (and mentally occupied/stimulated). It is these outcomes which appear to be both more important than work itself and more general in the respect that they can also be achieved through other means, such as volunteer work, childcare or organised trips or activities.

It was deemed that 'other people's understanding and acceptance' would determine how other people treated the individual with chronic pain, and that this would also impact on their self-respect and perceptions of their own

identity. A new label, 'Self-respect', was therefore created and elements from the themes 'Employment' and 'Other people's understanding and acceptance' can be thought of as determining either self-respect, identity, or the ability to remain physically and mentally active. This also suggests that 'remaining physically and mentally active' should be thought of as a new label, and it is logical that this would also be achieved through engaging in leisure and sporting activities.

On reflection it also appeared that 'mobility', 'sleep', 'medication', 'pain and discomfort', and to some extent 'mood', appeared to fit together and all determine an individual's physical and mental well-being; mood to the extent that this included depression.

'Individual understanding and acceptance' was a label given to questions that patients had regarding the cause of their pain (or the initial 'trigger'), along with their own efforts to come to terms with their illness. While these questions and feelings are undoubtedly important, it was not deemed appropriate to include them in any final measure as there remains a poor scientific understanding of chronic pain and it may not be possible to provide all patients with satisfactory answers that they will accept. Further, the objective here is to assess the impact of treatments/medical interventions/social policies. It can be argued that much of the process of understanding and coming to terms with ill health is dependant on the experience that a patient has during the periods of consultation with medical professionals, rather than on the actual treatment they receive.

Finally, it was felt that apart from the feeling of doing something worthwhile, meeting people and keeping one's self busy, many activities, particularly social activities, holidays and sport are done largely for enjoyment. Although physical and mental well-being would cover depression, this is more of a 'clinical' term/concept. Many patients reported that their pain made them feel tired and frustrated, or just prevented them from taking enjoyment from activities that they would ordinarily have taken enjoyment from. Patients also reported worrying about the future, how they will cope and if their pain will get

worse. Worry, frustration, and possible envy of others in more fortunate circumstances may all contribute to depression, but this does not seem to be a certainty. Failing to experience enjoyment from whatever activity or stimuli would seem distinctly different to actually experiencing clinical depression. While depression may be caused by or exacerbate problems with personal health and circumstances in the medium to long-term, enjoyment is surely more closely linked to single events which may last for hours or even minutes, and is something which we more actively seek to achieve. We have seen from previous chapters that Sen and Clark both suggest including happiness within the capability set, it is enjoyment from life which was brought up in our focus groups and it is suggested that enjoyment is a more appropriate concept to include (Nussbaum and Sen 1993; Clark 2005).

Take for example a cycle ride or a walk. These were activities which the men in the 18-65 (employed) group reported wanting to take part in with their partners and family for enjoyment, but the enjoyment was spoiled by pain or discomfort and a feeling that their partner's enjoyment was also being diminished by their very presence. While continually experiencing such feelings of disappointment, frustration, guilt, and so on may lead to depression, it would not be sensible to expect people to simply choose not to be depressed, but we should expect people to be able to choose to do something which they enjoy and to experience this enjoyment. Enjoyment itself then appears to be a reasonable and important capability.

The process resulted in ten labels:

1. Self-Respect
2. Social interaction
3. Role of parent/grandparent
4. Support from family
5. Remaining physically and mentally active
6. Identity
7. Independence & Control
8. Relationships
9. Physical & mental wellbeing

10. Enjoyment

Finally, these labels were rewritten and interpreted as ten 'higher order capabilities' and the themes interpreted as other capabilities or functionings relating to, and indicating the level of achievement in, each of these ten areas (see Figure 9.6). It was appropriate to interpret the data in this way as the groups were asked about how their chronic pain restricted their freedom and ability to perform tasks and roles deemed necessary for a good quality of life. Figures 9.7 to 9.16 provide further illustrations of the ten capabilities.

Figure 9.6 Initial List of Ten Higher-Order Capabilities and Associated Functionings

<p>A. To be able to achieve self-respect.</p> <ul style="list-style-type: none">➤ Being treated with respect by others➤ Being honest about your condition and your feelings➤ Being believed➤ Not having to rely on others to help with everyday tasks➤ Doing something worthwhile (in the home, at work or through voluntary activities)➤ Earning a wage
<p>B. To have reasonable opportunity to enjoy social interaction.</p> <ul style="list-style-type: none">➤ Having equal and unrestricted access to employment➤ Being able to get to and fully take part in social events➤ Being able to visit friends➤ Being included in a social group➤ Having enough money to go out and socialise
<p>C. To be able to fulfil the role of parent/grandparent.</p> <ul style="list-style-type: none">➤ Being able to pick up, physically protect and hug children in your family➤ Being able to take part in activities/games/sports with children➤ Being able to show support for my children/grandchildren by watching them take part in sports/activities
<p>D. To have the ability to spend time with and feel well supported by your family.</p> <ul style="list-style-type: none">➤ Being able to travel to see and stay with family members➤ Being able to go on holiday with family➤ Having the support of your family
<p>E. To be able to remain both physically and mentally active</p> <ul style="list-style-type: none">➤ Being able to pursue any hobbies and interests (e.g. theatre, fishing, bowls) that you were able to do before your pain.➤ Being able to watch and/or play sport/activities (e.g. dancing, horse-riding, football)➤ Having equal and unrestricted access to employment

F. To be recognised as a person with an individual identity – not be seen as an illness.

- Being free of the tiredness and frustration associated with chronic pain
- Not having to ask for special treatment
- Not being treated differently by others
- Being able to talk about work, hobbies and interests
- Not having to cover up pain

G. To be able to be independent and to feel like you have control over your own life.

- Able to care for oneself (dress, shower, use the toilet)
- Able to drive and/or use public transport
- Able to walk to the local shop or bus stop
- Not having to base your routine around taking medication
- Being able to manage stairs
- Not relying on crutches/sticks/a wheelchair
- Being able to cook and manage the home

H. Not be constrained in your ability to experience a loving relationship.

- Meeting people (potential partners)
- Being independent – not needing help from other people
- Having disposable income
- Being able to enjoy physical affection
- Being able to fully take part in social activities and holidays with partners
- Having patience and not feeling frustration due to pain

I. To be able to achieve a good level of physical and mental well-being (similar to most people your age).

- Being free of pain (when sitting, standing, walking)
- Not be worried about the long-term effects of medication on health
- Free of negative effects of medication (e.g. tiredness, dizziness)
- Able to sleep well
- Able to exercise
- Free of depression

J. To be able to find enjoyment in life.

- Be free of tiredness and frustration
- Able to fully take part in work, sport, leisure, social activities
- Not feel ashamed or envious
- Not fear the future

Figure 9.7: A Summary and Illustration of the Capability: To be able to achieve self-respect

Respondents reported being treated with a lack of respect by others. Many respondents felt that, because the presence of pain is not easily visible, other people questioned whether their pain is genuine. Some worried that members of the public, neighbours, and colleagues think that they are 'putting on an act' or 'seeking attention'. There were several stories of how, when attempting to claim income support or other assistance, 'officials' were hostile and challenged the extent to which patients were debilitated by the pain. In one case, a respondent reported having given up even attempting to claim support after being called in to the benefits office three times in one week to be questioned on the legitimacy of his claim.

"... Until I gave them the doctor's thing. You know, it's the way they looked at me, because they didn't believe me. I filled out all the forms how you're meant to fill them out, but you put down something from chronic pain and that and they sort of query it" (M 18-65, B)

"I tell you. I get more respect if I'm in this chair, knowing that I can't stand or walk, than I do if I'm on my mobile. People will *not* respect me if I'm on my mobile. ... The thing of it is you see, with a mobile, people don't know that I'm not able to stand or walk. You see people on mobiles and they park outside a shop, and they get off, do their shopping, come out. And I think that's the reason." (F retired, C)

Because of the fear that people did not believe them and a belief that others simply had no understanding of their condition, and hence could not be empathetic, many respondents spoke of a battle to find a balance between being honest and 'selling the story' that they are suffering less than they actually are.

"My son' at university now and I notice when he rings up he says '[Sigh] Hi Mum, how are you? And I know from that sigh that he wants me to say 'I'm great, I haven't been ill for days, I feel fantastic, I've done X'. And I don't actually tell him the truth because I know he's getting quite bored with the truth... so erm I kind of fudge my way around it and make my life sound far more exciting than it actually is, so as not to bore him to death" (F 18-65, D)

"I always hate, when sort of say, someone says 'how are you?', I can't lie to 'em, I can't turn round say I'm absolutely bloody fine, when I'm not, you know. But you can't, you can't turn round and say ah, terrible!" (M 18-65, B)

Respondents often felt embarrassed at having to ask for help or for special treatment. Similarly, respondents worried that other people viewed them unfavourably when *they* were unable to help other people or play a productive role in society. Linked to this, and the hassle of claiming support, was the ability to earn a wage; to support one's self and one's family.

"You feel as though, er... [Interrupted]" (M employed, A)

"[Interjects] You're putting everybody out." (M employed, B)

"Er, yeah. Basically, yeah." (M employed, A)

Figure 9.7 Continued...

"In my sense ... you feel inadequate. Well I do, when my missus starts ... unloading the car, and you know, when kind of I walk into the house and sit down." (M employed, A)

"You feel like you're being attention seeking, but you're not... You're constantly having to ask for special treatment and I find that embarrassing" (F 18-65, D)

"Well, sometimes I feel embarrassed. When you ask people to help you." (F retired, B)

"You just feel as though you're letting everybody down. Well I do." (M employed, A)

"You're automatically type cast aren't you, sometimes? You know, you say you're not working and people think 'sponger'." (Partner of M 18-65, B)

The combination of these factors appeared to lead some respondents to question their worth as a person and to judge themselves as weak or hopeless.

"Why aren't I strong enough to cope with this shit pain? Am I weak?" (M 18-65, C)

"You just feel as though you want to hide away in a corner sometimes; when you're with a lot of people, you feel ashamed". (F retired, C)

"... you just feel hopeless, you *really* feel hopeless you do, but you just can't do anything." (M 18-65, B)

Figure 9.8: A Summary and Illustration of the Capability: To have reasonable opportunity to enjoy social interaction

Respondents spoke of how employment was a way of meeting new people and maintaining friendships. After being away from work for long periods, or in other cases having given up work, these friendships were difficult to sustain and the common bond and often the source of much conversation was lost. Consider the respondent quoted in Figure 9.4 and who spoke about giving up a “very social job” and subsequently losing contact with former colleagues.

Because of limited mobility and often an inability to sit or stand comfortably for long periods, respondents reported difficulties in taking part in social activities.

“...my social life is virtually nil; has become virtually nil. Even er going to a neighbour for a cup of tea, they all stand in the kitchen, whereas I can’t and I know that they do all come out to the dining room and sit down, it just makes it different, so I tend not to do that” (F 18-60, C)

“I think it’s easier not to go and have to explain for half an hour why you can’t do something, it’s so much easier just to say ‘I can’t be bothered I’ll stay at home’, it’s so much easier” (F 18-60 D)

The inconvenience and awkwardness of this can lead to a sort of self-withdrawal.

“It got to a stage, and Graham will back me up on this. It got to a stage where I just didn’t want to go out anymore” (F retired, C)

“Now I don’t want to, I don’t want to go out” (M18-65, B)

Or, it can mean that the invitations simply diminish or stop.

“Well, my friends, they started ringing, and I’d ring them, and they rang me when I came back, because I was in Australia last Christmas, and they rang me and I rang them a few times, and ‘oh yes, we’ll go out’, I haven’t heard from them since” (F 18-60, A)

“People phone you up and invite you out, but after a while they stop inviting you out cos they know you’re going to come up with some excuse, so your ability to meet people reduces over time I think” (M 18-65, C)

Many respondents spoke of feeling isolated.

“... it makes me, it erm, isolates me, I’m very isolated, I don’t see people very often, in fact someone came past my garden the other day and said ‘gosh I thought you’d moved cos I haven’t seen you for two years’, so because I’m not going out I’m very, very isolated.’ (F18-60, A)

Figure 9.9: A Summary and Illustration of the Capability: To be able to fulfil the role of parent/grandparent

In several of the focus groups the first comments made by respondents related to children and grandchildren. Often this related to the ability to pick up, hug and protect young children.

“Playing with the grandchildren is one thing... sitting on your lap, having cuddles with them” (F 18-60, A)

“Yes, I had to stop picking up my granddaughter, oh, she wasn’t even two, but I couldn’t pick her up, which is something I love doing, I love cuddling her, but I can’t do it now. I know she’s a lot older now, but you still can’t have them on your lap.” (F18-60, B)

“Do you think that affects your relationship with your grandchildren? (PK)

“Yes, when a little girl turns round and says ‘no, you can’t pick me up granny ‘cause of your back, that’s, that’s heart breaking, that really is. You should be able to do it” (F18-60, B)

“It’s always like mind your grandma, and if you’re jumping on the sofa, don’t jump on the sofa with grandma. And if you’re going to give her a hug, be gentle. And that sort of thing” (F 18-60, C)

“One of the things I’ve had with, er, with my kids, when they’re at school and that like that, sort of like the bullying aspect of it ... A couple of the children there were saying er, to my eldest boy, er, ‘oh, you’re Dad walks with sticks and that, or he’s in a wheelchair’. ... And he come home from school and he was *really* upset about it. ... And then with the other kids sort of kicking him and taking the micky out of him and that, that is, that is one thing what I feel, sort of, I’ve lost. You know, I can’t be there for them. You know, if they fall over or something ... I can’t get down to pick them up, and that’s, that’s the bit that hurts me, knowing I can’t do that for them ... I feel that a father should be able to do that for them ... and that’s one thing I really miss.” (M 18-65, B)

“If my grandchildren come to sit on my knee I put my leg under the table” (F retired, E)

“Again, when I can’t pick her up or comfort her. My husband seems to have a better relationship with her in that way, in that if she hurts herself she’ll run to him for a hug, rather than me because I can’t pick her up, whereas she knows that he can... You feel a bit left out really.” (F employed)

Not being able to play with and take part in activities with children was also an issue.

“... Even though he’s grown up with me being how I am, he still can’t comprehend *why* I can’t play football with him, you know.” (M 18-65, B)

“When I take the grandchildren out now I can’t play with them like I used to.” (M retired, B)

Some respondents also felt that their pain made it difficult for them to go and support children when they were taking part in activities themselves. For example, one male respondent reported how just walking across the pitches to where his son is playing football can mean that he has extreme pain.

Figure 9.10: A Summary and Illustration of the Capability: To have the ability to spend time with and feel well supported by your family

Respondents appeared to expect a greater level of understanding and support from family members than from other friends or colleagues and appeared to find it more hurtful when this was not forthcoming. The first quote below is from a respondent who complained that his father had never been to visit him in hospital, despite a period of frequent stays.

“My own father don’t believe me” (M 18-65, B)

“I find it particularly hard when it’s family. And it’s something that you perhaps don’t particularly want to do anyway, then they think you’re using it as an excuse. Do you know what I mean? So, like, family have asked you to help out and do something, and you say well actually I can’t, I’m in a lot of pain. And they think ‘oh yeah, of course she is, she’s just trying to get out of it.’” (F employed)

“You tell them, but they don’t know how you feel. Your husband doesn’t know how you feel. Your sister doesn’t know how you feel. They can’t do, because they can’t feel your pain.” (F retired, A)

“I don’t know if anyone else feels guilty, I have a huge guilt complex that I worry that I put all this stuff on to my husband, I put all this stress on to my Mum, cos she says ‘what did I do to deserve, you know, what did I do to have two daughters that are so poorly’ and things like that, so I think if I wasn’t poorly she’d only have one...” (F 18-60, C)

Many respondents spoke of their difficulty at visiting or holidaying with family.

“I mean, I’ve got a sister who’s now moved to [unclear]. She’s now told me that she’s got two steps, three steps in her bungalow, so no way am I going to be able to get in there. So where, do you go and see her, or not? Do you sit outside and have a cup of tea?” (F retired, C)

“The other thing is things like holidays and family outings. Again, you have to sort of plan them around when is likely to be a good time, pain wise.” (F employed)

Figure 9.11: A Summary and Illustration of the Capability: To be able to remain both physically and mentally active

Most respondents spoke of having to give up certain hobbies and interests, or experiencing greater difficulty in undertaking them.

"I find that I can't arrange to do anything, because I never know how I'm going to be, you know. I still try to stay in contact with some of my friends from the fishing club, what I used to, sort of help to run and that. I can't, they have monthly meetings and I still like to try to go, to go to them, but I can't say from one month to the next month that I'm going to be there, 'cause I never know how I'm going to feel." (M 18-65, B)

"I used to love my fishing and all that, but as its got worse and worse and worse, I've sort of, you know I used to go every weekend, but now I can only go if my brother goes, if he comes to pick me up, gets me all my stuff ready and all that, and I'll sit there with my sandwiches and wait for the float to go, and that's it" (M 18-65, A)

It seems that, especially in cases where respondents have had to give up work or are retired, hobbies are important to respondents in order to pass the time and prevent boredom.

"...keep you occupied, and your mind. The reason I knit and cross-stitch is to, I've got to do something. Otherwise you'd go crazy." (F retired, C)

"Work is important. Just to go out and do things that aren't mind numbing if you know what I mean." (F employed)

"What upsets me is that I'm unable to work. I'm only 39 years old, I think to myself, is that my life already. It makes me depressed." (Male, 18-65, B)

Other, more physically demanding activities were also important to pass the time, for enjoyment and in order to remain healthy. One male respondent reported how his wife and children had all bought mountain bikes whereas he had not because he felt he would have difficulties and discomfort in riding one.

"I mean we used to, we used to go long distance walking. We've done the Peddars Way, The Norfolk Coast Path, Angles Way, everything, but now, I mean?" (F retired, C)

Figure 9.12: A Summary and Illustration of the Capability: To be recognised as a person with an individual identity – not be seen as an illness

Respondents felt that the constant pain and the resulting tiredness, frustration and, in some cases, depression altered their personality.

“It destroys your personality, destroys it I think” (M18-65, C)

“The pain makes you grumpy, and so you’re looking for reasons to be grumpy.” (F employed)

“When you’ve got children, they know what’s matter with you. And they sort of get the blunt end of it sometimes” (F retired, B)

“Your frustrations? (PK)

“Yes” (F, retired, B)

In addition, respondents spoke of how they were often viewed as a person with an illness, and how it felt sometimes how the illness had become their identity. Often, having to ask for help, or for special treatment exasperated this.

“If you think, like to think of yourself as an interesting person, you don’t want to become an illness, so the first thing people say to me is ‘how are you today?’, not ‘what have you been doing?’, er, it’s not about what I do anymore (F 18-60, C)

“I notice it more in the wheelchair. People don’t talk to me, they talk to who’s pushing me, so I’m just the person sitting there, so I’m obviously not capable of any thought or decision” (F retired, C)

“You don’t want to be seen as constantly complaining about something.” (F employed)

Because of having to give up work or hobbies, respondents felt that little happened in their lives that they felt they wanted to discuss in conversation. It was also reported that other people tended to focus on the pain/illness when making conversation. Despite this tendency for people to ask, respondents spoke of how they felt that other people were bored by their pain and confused by the fact that they were experiencing no improvement. As a result, respondents often spoke of covering up their pain.

“... I can’t cover it up all the time. Sometimes I can cover it and make it look like I’m not in pain, but it’s an awful lot of energy” ... “When people ask if you’re alright, at some point they want to hear a positive response don’t they, and if they don’t, they don’t know how to deal with it.” (M 18-65, C)

“... it’s boring for you to keep saying ‘actually I don’t feel that good’, and I think it’s boring for other people. So I think generally you don’t really say how you feel” (F 18-60 C)

Figure 9.13: A summary and Illustration of the Capability: To be able to feel independent and to feel like you have control over your own life

“It is losing your independence isn't it, if someone's got to do everything for you.” (F retired, B)

At home, respondents spoke of how their pain limited their ability to cook, look after the home and to dress, shower and take a bath. Some listed adaptations to the home and aids such as crutches, a walk in shower, a seat in the shower and blocks to make chairs higher.

“Well, I wouldn't be able to get in the bath” (F Retired, D)

“Some housework you can't do can you.” (F retired A)

“I can't pick up saucepans and this and that, I live on pizzas” (M 18-65, A).

“Now, I have to instruct someone when to take something out the freezer, erm, and, it's very frustrating, trying to explain to somebody how to do something in the kitchen...” (F 18-60, B)

There were also limitations in terms of being able to leave the home, travel, shop and so on.

“I can't walk to the doctors, I have to get a taxi” (M 18-65, A)

“... say you want some milk from the shop, that's only just down the road, but I can't walk and get it...” (F retired, B)

“I'm not able to drive anymore. Whereas I used to drive” (F retired, C)

“Is it due to pain that you're unable to drive?” (PK)

“Well, I can't put any pressure on my feet, and I've got arthritis and that in this arm...” (F retired, C)

“...Can I ask, is driving something that you regard as being important?” (PK)

“Oh yes. Well, its independence isn't it. I'm so reliable on Graham that he has to be with me really all of the time” (F retired, C)

Some respondents spoke of having to base their routine around their pain or their medication.

“This morning, I got up - 5 o'clock - I took my first pain killers, went back to bed again so that I was ready to get up to have my shower at half past six, or else, by the time you start taking them they haven't taken effect and you're trying to move around. So, yeah, you've got to think ahead...” (F 18-60, A)

Figure 9.14: A Summary and Illustration of the Capability: Not be constrained in your ability to experience a loving relationship

Of those respondents who were single, some spoke of how their pain made it unlikely that they would meet a potential partner and then build a relationship.

“I’m on my own, sometimes very lonely, erm, being in pain it feels like you’re on your own all the time, and not being able to build a relationship with somebody because, the chronic pain is always there, it makes you grumpy, it makes you depressed.” (M 18-65, C)

“You know how you said there about keeping a relationship. You know, when you can’t work, and we’re on, we’re all on income support, or DLA or whatever. I’ve never got no money to take a woman out for a meal or something” (M 18-65, A)

Those who were in a relationship spoke of how their pain interfered with the physical relationship they had with their partner.

“Now, I’m 58, 59 sorry, and up ‘til kind of, up ‘til trapping this nerve, I had a very, very satisfying and happy sex life. But even that’s kind of dwindled down. Because, as I said, you’re in damn agony.” (M employed, A)

“My sex life has suffered, because my partner doesn’t seem to want to touch me, it’s as if, if he touches me he’s actually going to hurt me, but also he sees me in a completely different way now; I’m not the girl that he kind of fell in love with and my personality has obviously changed because, because of the pain, because of what the pain does to you. And so therefore he doesn’t see me as that person anymore... He certainly doesn’t see me as being sexy anymore ... he sees himself as a carer... I do sympathise with him for that because I know it must be difficult for him, but it’s also very difficult for me” (F 18-60, D)

“My husband will be home in December, and the first thing, he’ll grab hold of me, round the back, and he forgets and squeezes too hard, and how can you stand and groan when it’s the first time you’ve seen him in three months” (F 18-60, A)

“I have to have pillows all around me. So, really and truly, although I share a bed with my husband, still, I may just as well rather be on my own... And so it doesn’t help your marriage either does it? ... And then if he comes near me to cuddle me I’m on edge.” (F retired, A)

Most respondents who were in a relationship also spoke of how they feel guilty that because of them (or rather their pain) their partner is ‘missing out’.

“It puts a lot of pressure on our relationship, because it’s hard for him (a) seeing me in pain quite a lot, and (b) because, again, he can’t plan ahead, like I said, any plans and things can be altered at the last minute. I think he gets quite, sort of, frustrated about it. And that makes it quite difficult.” (F employed)

“I feel like I’ve let my wife down this last holiday, because there’s certain things that she’s wanted to do that, oh, if you want to do it love you go and do it, but there’s no way I’m even attempting that.” (M employed, A)

Figure 9.15: A Summary and Illustration of the Capability: To be able to achieve a good level of physical and mental well-being

The underlying causes of pain as well as symptoms and severity varied greatly between respondents, but the quotes below give some illustration of the problem:

“I’m in pain 24-7, whether I’m laying down, in the bath or hanging from the lightshade, I’m in bloody pain, you know” (M 18-65, B)

“OK, so we’re on discomfort. She’s alright when she’s lying down. When I’m lying down I’m terrible.” (F retired, A)

“I take morphine, but I try not to take it too often, and working through the pain, you’re debilitated, you’re sleepy after you’ve taken your pills. You’re exhausted. You’ve not done anything, you’ve just sat on a chair, but you’re exhausted” (F 18-60, C)

“Pain is very tiring” (F 18-60, B)

“Yes, it is very tiring” (F 18-60, C)

Pain effected sleep and the ability to exercise (stay healthy):

“I think a good night’s sleep is vital. And, er, without the medication, I don’t get that” (M retired, A)

“I have a sleeping tablet because I can’t sleep, because if I move in my sleep I’m awake, that’s it” (M 18-65, A)

“...dancing and walking and all those sorts of things. You were getting exercise. The next thing is you’ve got to watch what you eat because you don’t want to put on weight...” (F retired, A)

“How do you exercise when it’s kind of giving you agony?” (M employed, A)

Medication often resulted in unpleasant side-effects.

“I keep trying stuff [medication] but it’s not nice, there is one out there waiting with my name on it. So for me, lack of medication is a problem, the medication that I have tried has been horrible, has done horrible things to me, so that’s a problem...” [F 18-60, B]

“A lot of the medication I take has ghastly side effects” (F 18-65, D)

“I find that with the morphine that I’m on, every, every about a month, two months, I’ll have a real bad couple of day when I’m just freezing cold and I can’t do absolutely anything, you know and I’ll just be in bed and that...” (M 18-65, B)

“... In the end, your body is so stocked up with all of this rubbish inside of it that you don’t function properly. And, and you don’t know who you are even. You do things and you don’t know how you did it, but you just did it. Until, someone wakes you up and say’s for God’s sake stop taking all these blinking rubbish tablets...” (F retired, A)

Figure 9.15 Continued...

Many of the respondents also reported experiencing some degree of depression.

"I get, I started getting depression... it's like yesterday, I didn't have a wash, I didn't have a shave, I didn't get up, I didn't even unlock the door, and that was it" (Male 18-65 A)

"I think depression is hand in hand with chronic pain" (Male 18-65, C)

"It leads to depression and that sort of thing..." (F employed)

Figure 9.16: A Summary and Illustration of the Capability: To be able to find enjoyment in life

Depression was not experienced by all respondents, and depression appeared to be correlated with the severity of the pain and the amount of time spent living with pain. What seemed to be more common was the presence of good and bad days. Respondents spoke of having days on which they generally felt positive, and days when, due to tiredness, frustration, or pain severity, they felt negative. Respondents also spoke of how, more consistently, there were activities that they used to enjoy, but no longer took enjoyment from.

"I think frustration is a big thing" (F retired, E)

"The negativity is what kind of mood you're in, and whether you're actually, whether you actually feel like you've got the energy, the mental energy, not the physical energy to climb over this mountain of problems that you know you've got in front of you" (F 18-60, D)

"It's different days isn't it? Some days, some days I feel fine, up here, and if I feel fine up top I'll have a better day. But if I wake up in the morning and I don't feel right up here, you know the day is going to be worse" (M 18-65, B)

"You lose interest in, er, in shopping really. You know, it just, becomes a grudge" (F retired, C)

"You've lost the enjoyment?" (PK)

"That's right. There's no browsing anymore." (F retired, C)

"I can't dance anymore. And I did love dancing... I went to a wedding in September and that was the first time that it actually hit me. They had beautiful music playing, and I couldn't get up and dance. And I could have cried my eyes out." (F retired, A)

"If someone had told me ten years ago that I was going to be irritated because my partner had to do the ironing I'd have said no way" (F 18-60, D)

"You're missing out on what you used to do" (F 18-60, B)

Figure 9.16 Continued...

Witnessing other people having fun, or just doing activities that the respondents were unable to do themselves, appeared to make their situation more difficult.

“Well you sort of get jealous of other people. Is it jealousy? I don’t know what the right word is? Envious. Erm, seeing people, even when I sit and they’re walking past the window. And you think ‘oh God, that must be lovely to be able to get up and walk down the road.” (F retired, D)

“You do feel resentful. Sort of why have I got it bad, and that sort of thing.”

“When you’re with a lot of people, that’s when you realise what you can’t do” (F retired A)

Finally, concerns for the future were another burden for some respondents. Concerns mainly stemmed from the knowledge that their condition was likely to worsen, and related to finances, relationships, and their ability to cope.

“Time after time after time you get no answer, and so in a sense you’re losing your hope all the time...” (M 18-65, C)

“My main concern is if anything happened to my husband, he’s a bit older than me, he’s 14 years older than me, yeah, he’s coming up 64 I think, and he’s fit and healthy and everything, but he is, er, everything, he’s always been my world anyway, but now he is my physical world as well as my emotional world because he does everything. So, that is, something that is something that is always in the back of my mind that, he’ll go out and do something, it’s always ‘be careful’, ‘drive careful’...” (F 18-60, C).

“Sometimes you do feel very down. And you think, I’ve got this forever basically.” (F employed)

“It is worrying. It’s an extreme worry, having the children and not being able to work. I’m only 47...” (M employed, B)

9.5 Individual Interviews

At the end of each of the focus groups the participants were asked if they were prepared to be approached to take part in a further individual interview; the purpose of which would be to review the list of capabilities and functionings. Those who agreed to be contacted were asked to sign a consent form (presented in the same format as that in Appendix A). Once the analysis was completed and the list of key capabilities identified from the data, one or two participants from each group were contacted by telephone to request an interview. Those who agreed to be interviewed were offered the

choice of conducting the interview at the participant's home, or at the James Paget Hospital.

Participants had discussed their chronic pain at focus groups which contained specific segments of the sample group. The intention was to make the groups as homogenous as possible, in turn promoting free and more relaxed discussion (Morgan 1997). A further reason for segmenting the overall sample was that it would be more obvious if a particular theme was only relevant to one 'type' of participant. Although it appeared that the main themes identified were relevant to all groups, there was a risk that in generalising, summarising and re-phrasing the main themes the initial meaning (as intended by the participant) would be lost. The purpose of the interviews was therefore to check that important themes discussed in the focus groups had not been missed or wrongly interpreted, and that after the main themes had been described in different language by the researcher, the meaning of the themes was clear to participants.

It proved difficult to contact some participants and because of this, and the fact that participation in the focus groups had been lower than expected, it was not possible to interview a participant from every segment. Interviews were conducted between 22nd January and 14th February 2007. Three were held at the patient's home and three were held at the hospital, with the location being chosen by the patient.

The interviews began with an explanation by the interviewer (myself) that the list of capabilities that was about to be presented to the participant for discussion in the interview was directly based on discussion from the focus groups. Participants were reminded that in the groups they had been asked how their pain restricted their freedom and ability to achieve important tasks or roles, and that the moderator had then asked them to further explain how this had an impact on their quality of life and what aspects of quality of life were affected. It was explained that in the interviews things would be considered from the other direction; they were now going to be asked to consider a list of 10 component aspects of a good quality of life, and listed

under each of these was a number of examples of activities/freedoms that were symptomatic of the achievement of each of the 10 aspects of a good quality of life. The first capability (“to be able to achieve self-respect”) was used as an example. So, it was stated that having self-respect was thought to be important in order to achieve a good quality of life. It was then stated that things such as “being treated with respect by others” and “being believed” were considered to be important in order to achieve self-respect.

Once the list of capabilities and functionings had been introduced to the participant the full list was given to the participant to read through. Next, the interviewer read out each capability and each functioning in turn as a statement and asked the participant a number of questions about each one. Participants were asked if the meaning of each ‘statement’ was clear and if they agreed that each was important. When all of the capabilities had been discussed the participant was asked if there was anything that they could think of which had been missed off the list.

An audio recording was made during the interviews and the interviewer made brief notes directly on to their own copy of the list of capabilities and functionings.

9.5.1 Feedback from interviews

Six patients (and one partner of a patient, who was present at the time of the interview) were given the list of ten capabilities and associated functionings and asked for their thoughts/opinions on: (i) if they considered the capabilities to be important, (ii) if the meaning was clear to them, and (iii) if anything important had been missed.

Respondents generally felt that how they were treated by other people was important; one respondent who used a wheelchair told of how people conducted conversations above her head, two respondents said that they attempted to hide the fact that they were in pain, another worried about being thought of as a burden.

It was clear that some respondents did not differentiate between social interaction and time with family, and indeed respondents reported going on holiday with family and some said that rather than visiting friends they would tend to (or like to) visit family. Three respondents felt that being able to offer younger members of their family financial support was also something that is important. Three of the respondents felt that as well as receiving support from family it was also important that their family believed that their pain was real – that their family accepted their illness. There were reports in the groups of negativity from family members, as well as denial.

There appeared to be something of a divide between respondents. Some felt that it was down to them to try and ‘carry on as normal’ and to avoid ‘moaning’ about their illness, in order to be treated normally by others. One respondent, for example, spoke of the difficulty associated with ‘trying to make yourself interesting’. Other respondents expressed the opinion that it is other people who need to be more accepting.

Independence and control was an area that respondents clearly felt was important and many offered their own examples of things that they had to plan in advance or seek help with, as well as the frustration of not being able to do things they had been able to do in the past.

Respondents generally seemed puzzled or expressed doubts about including some additional concept of independence under the heading of relationships and so this was removed.

As an additional task, respondents were asked to pick out and order the three capabilities that they regarded as having the greatest importance for a good quality of life, and then to position these on a visual analogue scale of 0 (Not at all important) to 1 (Extremely important). One female participant in the older group was accompanied to the interview by her husband, who also simultaneously completed this final part of the interview.

Four capabilities were selected as being the most important: physical and mental well-being, the role of parent/grandparent, independence, and physically and mentally active (see Table 9.3). Using a system of points, it is possible to rank the capabilities. Here, three points will be awarded if a capability was deemed by a respondent to be the most important, two if a capability was deemed to be the second most important, and one if a capability was deemed to be the third most important. Results are presented in Table 9.3. Because the partner of one respondent also completed this exercise, results are presented which include his responses and exclude his responses; no difference is made to the rank ordering of the capabilities if his responses are included or not.

Table 9.3: A Ranking of the Ten Capabilities (highest to lowest)

Capability	Score	Score (excluding partner)
Independence & Control	12	10
Physical & mental well-being	9	6
Role of parent/ grandparent	6	6
Physically & mentally active	5	5
Self-respect	3	3
Enjoyment	3	3
Loving relationship	3	2
Social interaction	1	1
Identity	0	0
Family	0	0

Alternatively, the capabilities can be ranked according to scores from the visual analogue scale, see Table 9.4.

Table 9.4: Ranking of the Ten Capabilities based on VAS Scores (highest to lowest)

Capability	VAS scores	VAS Scores (excluding partner)
Independence & control	5.0	4.2
Physically & mentally active	3.2	2.4
Physical & mental well-being	2.8	2.0
Role of parent/ grandparent	1.9	1.9
Self-respect	1.4	1.4
Loving relationship	1.0	1.0
Enjoyment	-	-
Social interaction	-	-
Identity	-	-
Family	-	-

Table 9.3 varies significantly from Table 9.4 because one respondent who was able to identify three capabilities as being the most important did not attempt to place them at points along the visual analogue scale and because one respondent changed his selection of the three most important capabilities between the first part of the exercise and completion of the visual analogue scale.

Following the interviews it was decided that ‘To have the ability to spend time with and feel well supported by your family’ should be dropped as a higher-order capability, as it appeared to contain functionings either covered under the area of social interaction, or which related to self-respect. It should be noted that the intention of the interviews was not to shorten the list of capabilities, but rather to check that the themes and labels drawn from the focus groups had been correctly interpreted and were described clearly. A decision was made not to cut anything out which had been discussed by the respondents in the focus groups unless the evidence from the interviews was that it was obviously redundant or had been wrongly interpreted. In other words, respondents essentially had the final say over what was included and there was no external interference with this. The final list of nine key capabilities is contained in Figure 9.17.

Figure 9.17: Final List of Higher-Order Capabilities

<p>1. To be able to have self-respect</p> <ul style="list-style-type: none"> ▪ Being treated with respect by others ▪ Being able to be honest about one’s pain ▪ Being believed when talking about one’s pain ▪ Feeling that one’s core family have accepted (come to terms with the fact that) one has chronic pain ▪ Not having to rely on others for help with basic, everyday tasks ▪ Being able to do things considered to be worthwhile and productive ▪ Being able to earn a wage
<p>2. To be able to enjoy social interaction</p> <ul style="list-style-type: none"> ▪ Having the opportunity to meet new people through employment ▪ Being able to go out and socialise ▪ Being able to visit friends & family ▪ Feeling included as part of a social group ▪ Having enough money to go out and socialise

<p>3. Being able to fulfil the role of parent/ grandparent</p> <ul style="list-style-type: none"> ▪ Being able to pick up, physically protect and hug young children ▪ Being able to take part in activities/games/sports with children ▪ Being able to support one's family financially
<p>4. Being able to remain physically & mentally active</p> <ul style="list-style-type: none"> ▪ Being able to pursue hobbies and interests ▪ Being able to do more physically demanding leisure activities ▪ Having the opportunity to do paid work
<p>5. Being able to have a positive & individual identity</p> <ul style="list-style-type: none"> ▪ Being free of the tiredness and frustration associated with chronic pain ▪ Not having to rely on others for special treatment ▪ Not being treated differently by others ▪ Being able to engage in conversation about work, hobbies or interests ▪ Not having to hide pain
<p>6. Being able to be independent/ have control</p> <ul style="list-style-type: none"> ▪ Able to care for oneself (dress, shower, use the toilet) ▪ Able to drive and/or use public transport ▪ Able to walk to the local shop or bus stop ▪ Not having to base your routine around taking medication ▪ Being able to manage stairs ▪ Being able to plan things in advance ▪ Being able to cook and manage the home
<p>7. Being able to participate in a loving relationship</p> <ul style="list-style-type: none"> ▪ Having the opportunity to meet potential partners ▪ Having enough money to go out and enjoy time with a potential partner/ date ▪ Not feeling bad about a partner missing out ▪ Having patience and showing emotional support ▪ Being able to enjoy physical affection
<p>8. Being able to enjoy good physical & mental well-being</p> <ul style="list-style-type: none"> ▪ Being free of pain (when sitting, standing & walking) ▪ Not be worried about the effects of current medication on future health ▪ Be free of any negative side-effects from medication ▪ Be able to sleep well ▪ Be able to exercise (in order to maintain or promote fitness/well-being) ▪ Be free of depression
<p>9. Being able to take enjoyment from life</p> <ul style="list-style-type: none"> ▪ Being free of tiredness and frustration ▪ Being able to take enjoyment from life ▪ To not feel ashamed, or envious of others ▪ To not fear the future

9.5.2 Similarities with lists of Capabilities developed elsewhere

The two capabilities Social Interaction and Relationships, are similar to that of 'Attachment' in the study by Grewal, Lewis *et al.* (2006) and Coast *et al.*

(Coast, Peters et al. 2008). Clearly, the two Capabilities 'Enjoyment' and 'Independence & Control' are virtually identical to those identified by Coast *et al.* (Enjoyment, and Control). The capability 'Role', defined by Grewal *et al.* as: "doing something that makes you feel valued", is similar to the functioning 'Being able to do things considered to be worthwhile and productive', which has here been included here as contributing to self-respect. The capability 'Security' (thinking about the future without concern) is incorporated here under Enjoyment (not fearing the future).

It can therefore be seen that the five capabilities identified by Coast *et al.* are similar to the capabilities identified not just from the older groups within this study, but from across all of the six groups. It should, however, be noted that here the respondents all had a serious health problem and it is unclear to what extent a younger, healthy individual would, for example, have significant concerns about the future, at least in terms of health. Of course, the broader the definition of the capabilities the more they will apply to a broader range of people; it is quite possible that a younger person will worry about getting a job, about accumulating debt, and so on.

There are also some similarities between the higher-order capabilities on our list and those proposed by Nussbaum. Nussbaum (2003) includes on her list the capabilities (p41-42):

- (2) Bodily Health: Being able to have good health, including reproductive health; to be adequately nourished; to have adequate shelter
- (5) Emotions: Being able to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; in general, to love, to grieve, to experience longing, gratitude, and justified anger. Not having one's emotional development blighted by fear and anxiety.
- (7) Affiliation: ... B. Having the social bases of self-respect and non-humiliation; being able to be treated as a dignified being whose worth is equal to that of others.
- (9) Play: Being able to laugh, to play, to enjoy recreational activities.

Nussbaum's capabilities and their related descriptions are much more relevant to a developing country context, but there is nonetheless some clear overlap. It is this overlap and similarity between the capabilities developed through participatory methods with chronic pain patients and with older people, as well as with the list drawn up by Nussbaum which suggests that there is scope to develop fairly broad and generic capability based instruments to allow comparisons between different projects, interventions or between different sub-groups within a broader population. This is not to say that the need to develop new lists (or capability sets) for new contexts and cultures will ever be eradicated, but there are promising signs that a generic list, for example for the general UK population, would be adequately relevant to diverse groups and would therefore prove useful in evaluating and comparing whole ranges of projects and interventions.

9.6 Questionnaire Design

Following the interviews the number of higher-order capabilities was reduced from 10 to nine and a number of amendments were made to the wording of the other capabilities/functionings. The higher-order capabilities were summarised in one or two words, which were used to form the section headings for the questionnaire and the other capabilities/functionings were used as questions, under each of which was listed four possible answers/responses. For example:

Self-Respect

I feel that I am treated with respect by others:

- All of the time
- Most of the time
- Rarely
- Never

9.6.1 Other Material Contained in the questionnaire

One of the original objectives of this PhD was to use the Capability Questionnaire alongside data on costs to perform an economic evaluation of a Sativex, a new cannabinoid-based treatment for the relief of chronic pain. Because of this, questions were also included which were originally developed and used by Nicola Cooper (2000) to obtain data on the costs of rheumatoid arthritis. Because of the nature of the questions taken from Cooper's work, it was necessary to develop an Initial Questionnaire and a separate Follow-up Questionnaire, although the capability-based questions in each are identical.

The Initial Questionnaire was therefore designed to record three types of information. The first section recorded the respondent's personal details. Second, questions which were required for an assessment of financial costs incurred primarily by the respondent, together with their friends and family. Third, the impact of chronic pain on a respondent's quality of life assessed using the capability-based questions.

Unfortunately, due to the very low numbers of patients enrolled on the clinical trial it wasn't possible to gather sufficient data to conduct an economic evaluation. It also became apparent that the development of the capability instrument was in itself a major undertaking of significant academic interest. Although the questionnaires distributed in both pilot stages included sections aimed at gathering data on costs this data was not used within the project. The first sets of questions did, however, provide some useful information on the demographics of patients along with their work and family status.

9.7 Pilot Stage

Invitation letters and Patient information sheets (see Appendix B) were sent to a total of 32 patients who were due to attend appointments at the Pain Management Clinic at the James Paget NHS Hospital between 29th June and 6th July 2007. The letters were sent out one week before the date of the

patient's appointment. The intention was to recruit at least 12 patients. It was felt that this number of respondents would be adequate as the intention here was to observe respondents completing the questionnaire, time the process, and hold an in-depth discussion with them about their reaction to the questionnaire. This process was labour intensive and the intent was not to collect data *per se*, as this was the role of a larger pilot stage planned to follow. Instead, the intention was to gauge reaction to a lengthy questionnaire, to see if respondents would tolerate being presented with such a high number of questions and if they could complete them all within a reasonable time period. It was also an opportunity to gauge reaction (on a small scale) to questions phrased in terms of capability before many more copies of the questionnaire were produced and handed out to larger numbers of patients.

Upon attending their appointment patients were asked if they would like to take part in the study, and if so, they were asked to sign a Consent Form (see Appendix B). Patients were then asked to complete the Initial Capability-based Questionnaire (see Appendix C). Alongside the Capability-based Questionnaire, respondents completed the EQ-5D and a comparison of the two sets of results is presented in chapter 11. The time taken by patients to complete the questionnaire was recorded and patients were asked a number of questions, which are listed in Figure 9.18. Questions in italics were completed by the interviewer.

Figure 9.18: Questions contained on the Feedback Form for Pilot Interviews

1. Did you find that the instructions on how to complete the questionnaire were clear and easy to understand?
2. Did you find the instructions useful?
3. Did you find that the questions were easy to understand?
4. Did you find that any of the questions were difficult to answer?
- Why?
5. Was it clear how you should indicate your responses to each of the questions?
6. Did you find that any of the possible answers/ options were very similar? -
Were there any questions for which you did or could have marked more than one answer/ option?
7. Were there any questions for which none of the possible answers/ options seemed appropriate?
8. Have you any other comments or suggestions that you would like to add?
9. *Time taken to complete*
10. *Any comments/ observations by the researcher?*

The pilot stage was designed to address the following points (Fink 1995, p86):

- Are the instructions for completing the questionnaire clearly written?
- Are questions easy to understand?
- Do respondents know how to indicate responses?
- Are the response choices mutually exclusive?
- Are the response choices exhaustive?
- Do respondents have any suggestions regarding the addition or deletion of questions, the clarification of instructions, or improvements in format?

Table 9.5 contains a summary of the characteristics of respondents from the pilot phase, and Table 9.6 contains individual details of the characteristics of respondents.

Table 9.5: Summary of Respondent Characteristics:

Total Participants	12
Total: Female (18 – 59)	4
Total: Female (60+)	1
Total: Male (18 – 64)	6
Total: Male (65+)	1
Average Age	49.58 years
Minimum time taken to complete	12 min
Maximum time taken to complete	29 min
Average time	21 min
Least number of questions missed	0
Highest number of questions missed	7
Number of full responses	8

Table 9.6: Respondent Characteristics

I. D.	Age	Gender	Work Status	Time taken to complete	Notes
P01	26	Female	Off-Sick	15 min	
P02	54	Male	Off-sick	24 min	
P03	75	Male	Retired	26 min	Large-print Version
P05	42	Male	Full-time Carer	20 min	Usually uses a green filter due to dyslexia
P06	52	Male	Off-sick	12 min	Large-print Version
P07	43	Male	Employed Full-Time	20 min	
P08	33	Female	Housewife	18 min	
P09	58	Female	Unemployed	20 min	
P10	67	Female	Retired	29 min	
P11	57	Male	Employed Full-time	23 min	
P12	36	Male	Unemployed	23 min	With many interruptions
P13	52	Female	Off-sick	21 min	

It can be seen that the recruitment of an even distribution of respondents over the four groups was not achieved; the distribution of respondents matches the distribution of patients who had appointments over that particular time period and who were therefore available to recruit. What is also noticeable is the very small number of respondents who were employed; this may be an accurate reflection of the fact that the patient group which are referred to the clinic have severe long-term pain and are therefore more likely to be off work,

it may also be a reflection of the relatively high rates of unemployment in and around the Great Yarmouth area¹¹.

In the relatively few cases where questions were missed in the Capability Section of the questionnaire, the main reason for respondents missing questions appears to have been the fact that the actual questionnaire was printed double-sided, so where questions were missed the respondent missed a whole page.

In one case a respondent indicated that a question relating to his ability to “cook and look after the home” was not applicable to him, as his wife performed these roles, and cooking etc was simply something that he would never choose to do. This is indicative of a more serious problem, namely that of respondents not understanding that it is the ability to perform tasks and roles which is important, not whether they actually do them; in other words respondents are missing the central principle of the Capability Approach. While there is an initial explanation stating: “Please do not miss any questions. If a question asks you about something which is not relevant to you then you should provide an answer based on whether you feel you could do what the question asks if you were in a position where you had to try”. The actual question reads: “I have been able to cook and look after the house...”, and especially as this particular question is towards the end of the questionnaire, respondents will consider it some time after reading the instructions, if these were read in the first place. It was therefore decided that some change to the wording of this and any similar questions should be considered (see list of amendments below).

An additional problem that was found is that respondents wished to tick more than one level in response to the questions; this is despite some levels seeming to be clearly mutually exclusive, such as: “I have been able to visit friends and family as often as I could before my pain and with no difficulty/I have been able to visit friends and family as often as I could before, but with

¹¹ Despite Great Yarmouth being a seasonal resort, the rate of unemployment (measured by the claimant count) for the Great Yarmouth Borough Council area for July 2007 was 3.8%, when for Great Britain as a whole the rate was 2.3% (www.great-yarmouth.gov.uk).

some difficulty”. A further example of the same respondent answering by ticking multiple responses is that they reported being able to care for themselves with: “no help, but with difficulty”, and having “needed some help and/or having great difficulty dressing, showering or using the toilet”. Although in this latter example, the first level could be amended to “some difficulty”, or even “some minor difficulty”, the most likely reason for the selection of multiple levels is the variable nature of pain. It is usually the case that pain patients report having both ‘good’ and ‘bad’ days and their ability to perform tasks will naturally vary depending on their immediate level of pain, discomfort, mood, etc.

One question that caused a great deal of confusion was that relating to relationships. Here, the question that respondents answer is supposed to be dependent upon their status in terms of either being single or in a relationship. There is instruction on what to do, as well as the presentation of the question being designed to highlight the need to select an appropriate question, it would seem counterintuitive for a respondent to be asked on the one hand about their ability to meet new partners, and on the other about the effect of their health on their current relationship; but several respondents responded to both.

Following the pilot phase of the study, the following amendments were made to the questionnaire and these are outlined in Table 9.7.

Table 9.7: Amendments made to Questionnaire:

Question	Section	Changed From:	To:
1B	3	"I am able to be honest about my pain"	"I am able to be honest about the full severity of my pain"
Multiple	3	"Over the past 3 weeks..."	"Over the past month..."
3A	3	"Over the past 3 weeks I have been able to (would have been able to) pick up, physically protect and hug young children"	"Over the past month I have been able to (or would have been able to) ..." [emphasis added]
3B	3	"Over the past 3 weeks, I have been able to (would have been able to) take part in games/ activities/ sports with children:"	"Over the past month, I have been able to (or would have been able to) ..."
3C	3	"Over the past 3 <i>months</i> , I have been (would have been) able to support my family financially and provide the financial security that I was able to provide before my pain:"	"Over the past 3 months , I have been (or would have been) ..."
5A	3	"Over the past 3 weeks, I have been free of tiredness and/or frustration"	Question deleted due to repetition
7A & 7B	3		Change to the presentation of the question

The instructions at the start of Section 3 were thought to be sufficiently clear and comprehensive, but the following phrases were further highlighted by underlining:

- Please do not miss any questions.
- For all of the questions you should choose only one option from the four listed

Further copies of the questionnaire were printed single sided, and at the end of each page there was a clear indication given that there were more questions still to answer.

9.8 Summary

Chronic pain patients were initially recruited through the Pain Management Clinic at the James Paget NHS Hospital, Norfolk, to attend one of a series of six focus group discussions, the purpose of which were to identify a list of key

capabilities, all of which are restricted in some way by chronic pain. The patients were allocated to a specific segment according to age, gender, and employment status. Each focus group was made up of one segment of the overall sample group.

The participation rate for the focus groups is a limitation of the study (16 patients attended the groups), although it remains an open question whether additional useful data would have been gathered from larger groups. The segment of the sample 'female and employed' was represented by just one participant. It was therefore necessary to consider conducting further recruitment and to hold a second focus group for this segment with a greater number of participants. The question to be considered was whether more useful data, and in particular if any new issues/limitations were likely to arise, from further group discussions. The interview with the female in employment took place several days after the five group discussions and it was noticeable from conducting the interview and comparing the notes how the only new issue to be raised was that of the ability to have children. It was felt that this was an issue which related specifically to this one participant, a female whose pain related to a gynaecological condition. The more 'fundamental' issue related to parenting, and this was a theme which had already been identified. This was a similar story for each of the themes.

While they may have been illustrated with a range of examples or specific considerations, the themes identified related to a few broader or more fundamental topics. It appeared as though a point had been reached where there was a 'saturation' of information presented by participants (in terms of this broader list of topics) and it was therefore felt unnecessary to extend the period of recruitment and hold additional focus groups for selected segments. It is acknowledged that had the groups been larger there would be more confidence that this saturation point had been reached.

Framework analysis was used to identify a list of higher-order capabilities from the raw data collected during the group discussions and this list of capabilities was taken back to participants from the groups, who were asked

to review the list in terms of whether the capabilities were clearly stated and important, and whether the list was complete. Changes were made and the list of key capabilities was then developed into the form of a questionnaire. The questionnaire was then ready to be piloted with a new group of chronic pain patients.

The decision not to transcribe the tapes verbatim could be seen as a limitation of the study. It is acknowledged by Ritchie and Lewis (2003, p220) that although it is most common to work with verbatim transcriptions, it is feasible to work with the 'raw' tapes, or simply observational notes. In this study both the audio tapes and observational notes were referred to. It was possible to identify themes from the audio tapes and to record their location using the tape counter. These themes were grouped under 'main themes'. It is acknowledged that where verbatim transcriptions may have proved more useful was at the stage of challenging the location of themes under a particular main theme and summarising and assigning meaning to these main themes. As Richard Smith and Angela Robinson had been at the focus groups and had their observational notes to refer to, they were able to have some input into this final step in the framework analysis.

Participants also had an opportunity to give feedback on the final list of capabilities and offer feedback on whether the discussion in the groups had been accurately interpreted.

The questionnaire was initially piloted with twelve new patients, also recruited through the Pain Clinic, which involved the respondent completing the questionnaire in the presence of the researcher, before being asked questions about the clarity of instructions, their understanding of the capabilities, the ease with which the questionnaire was completed and suggestions for improvement. A number of amendments were made following the initial pilot stage and the questionnaire was then ready to send out to a larger number of respondents. The next chapter outlines the next stage of piloting, from which data were collected and analysed.

CHAPTER 10: Study Phase II: Trialling a Capability Instrument

The purpose of this phase of the study was to test the questionnaire with a small sample of chronic pain sufferers who had not participated in earlier stages of its development. More specifically, the intention was to gather evidence on whether respondents understood the questions and how to answer them. It was also important to look at the item response rate, particularly with a relatively long questionnaire. Furthermore, the piloting gives an opportunity to explore, simply and intuitively, how responses on the capability questionnaire compare to those on the EQ-5D instrument and the Visual Analogue Scale (EQ VAS).

This chapter outlines the methods used to recruit chronic pain patients (section 10.1), the methods used to distribute and collect back questionnaires (section 10.2) and the characteristics of respondents (section 10.3). Section 9.4 explores the results, in terms of the item response rate for the three instruments, evidence indicating whether respondents understood the questions, and a basic comparison of the three sets of data. Section 10.5 concludes.

An option that was considered was to ask members of the general public to complete the questionnaire. Asking members of the general public to complete the questionnaire would have provided further information on how people react to questions phrased in terms of capability and would have shown if it was possible for 'healthy' respondents to achieve the highest possible level of capability. It would also have allowed comparison between pain patients and a more general sample of the population. It was felt, however, that with having the questions so clearly focussed on chronic pain, the questionnaire would have seemed irrelevant and confusing to respondents who do not suffer chronic pain.

10.1 Recruitment

A decision was made only to recruit new patients (i.e. patients attending the Clinic for the first time) to participate in the piloting. It was felt that these patients would be most likely to start new forms of treatment following their initial visit, and therefore we could reasonably expect to detect some change in their quality of life over the following twelve months. It was also important that patients had not been involved in the first phase of the study. Discussions with staff at the Pain Management Clinic suggested that approximately 100 new patients would be seen at the Clinic over the three week period in which we would be recruiting. A participation rate of approximately 50% was thought to be achievable, which meant a target of recruiting 50 respondents. A higher participation rate was expected than in the first (qualitative) phase of the study, as in the first phase patients were put off by the idea of having to speak within a group and no extra travel would be needed in order to complete the questionnaires, unlike the focus groups.

Patients attending the clinic for a first appointment between 6th September and 21st September 2007 and between 11th October and 18th October 2007 were sent an Invitation Letter and a Patient Information Sheet (presented in the same format as that in Appendix B). The reason that this gap was left within the recruitment was so that the mailing out of the three-month Follow-up Questionnaire did not fall within the Christmas and New Year period.

The inclusion criteria were:

- Patients were aged over 18 years. (Few patients under the age of 18 years attend the clinic, and if children had been included in the study then issues relating to child protection would have had to have been considered.)
- Only new patients were selected (attending their first appointment at the Clinic).
- Patients were selected who could read and understand written English.

When patients actually attended their appointment they were approached by the researcher and asked if they had received the Invitation Letter and Information Sheet, and if they would like to ask any questions about the study. They were then asked if they would like to participate and, if so, they were asked to sign a Consent Form (presented in the same format as that found in Appendix B). They then completed the Initial Questionnaires while at the clinic. The Initial Questionnaires consisted of the Initial Capability Questionnaire (Appendix C) and the standard EQ-5D instrument (with the EQ VAS).

Once again, this part of the study was approved by the Norfolk Research Ethics Committee.

10.2 Follow-Up Questionnaires

Follow-up Questionnaires were sent out to respondents in the post at three, six, nine, and then 12 months after the Initial Questionnaires were completed. They were sent out as two batches, reflecting the gap during the period of recruitment. The Follow-up Questionnaires consisted of the Follow-up version of the Capability Questionnaire (Appendix C) and the standard EQ-5D instrument (with the EQ VAS). Respondents were also provided with a pre-paid envelope.

If the completed questionnaires were not returned by the respondent within two weeks then a postcard was sent out as a reminder (Appendix B). If the completed questionnaires were still not returned one week after this then the questionnaires were re-sent; in some cases a new postal address for the respondent was provided by staff at the Pain Management clinic. On a small number of occasions, respondents were also contacted by telephone, in line with the proposal accepted by the Ethics Committee.

Upon returning the six-month and the 12-month Follow-up Questionnaires the respondent was sent a shopping voucher, each to the value of £5.

The responses from the questionnaires were entered into the SPSS Statistics Package. At this stage, values for the levels within questions had not been elicited, and so responses were entered as follows: 0-No problem, 1-Some (minor) problem/difficulty, 2-Major problems/difficulty, 3-Complete inability/worst possible level.

The method by which values were elicited for the capability states and higher-order capabilities is outlined in Chapter 10.

10.3 Participant Characteristics

Table 10.1 shows that the average age of respondents was 58. From Table 10.2 it can be seen that a majority of respondents were retired and of the remaining 53%, most (33%) were not in work, either due to unemployment or sickness. This fits with the relatively low mean monthly income of £1019.40 per month. Many left this question blank, perhaps because they did not regard benefit or pension payments as income, depending on their perceived definition of the term. To put the average income of our respondents in some context, nationally, in 2008, those in full-time employment, who were also in the bottom 10% of the earnings distribution, earned less than £1135.33 per month (ONS); the average monthly wage in the UK up to April of 2008 was £2075.67.

Table 10.1 Age and Income

	N	Minimum	Maximum	Mean
Age	49	26.00	87.00	57.59
Income	31	300.00	3000.00	1019.40

It may be the case that the relatively low income of the respondents is one factor that will restrict their quality of life and this will either be an additional factor to, or more likely a result of, their chronic pain. According to the Chronic Pain Policy Coalition, 25% of those diagnosed with chronic pain go on to lose their jobs, and pain is the second most common reason given for claiming incapacity benefit (CPPC).

Table 10.2 Employment

	Frequency	Percent
Part-Time	4	8.2
Full-Time	6	12.2
Off-Sick	12	24.5
Unemployed	4	8.2
Retired	23	46.9
Student	0	0
Housewife/ Husband	0	0
Full-time Carer	0	0
<i>Total</i>	<i>49</i>	<i>100</i>

There was a reasonable representation of both males (53.1%) and females (46.9%) within the respondent group. Within the general UK population there are slightly more females, particularly at older ages (>73), but for a convenience sample, the mix is a relatively close reflection of the general population. It is unclear what the 'usual' characteristics of the patient population are.

10.4 Results

Here the intention was to answer three questions: (i) what is the item response rate? Have the same questions consistently been missed? (ii) Is there evidence that following an initial appointment at the Pain Management Clinic the patients have shown improvement in quality of life? (iii) How similar are the results from the three different instruments?

10.4.1 Item Response Rate

Table 10.3 gives a general overview of the number of questionnaires returned (EQ-5D, EQ VAS and Capability) which were incomplete. Forty eight respondents completed the initial questionnaire, but the numbers which were returned then declined over the 12 month period; Table 10.3 reports the numbers of questionnaires that *were returned*, but which were incomplete.

Table 10.3: Number of Incomplete Responses

Stage	Number Returned	Capability Questionnaire		EQ-5D		EQ VAS	
		Number Incomplete	Percentage	Number Incomplete	Percentage	Number Incomplete	Percentage
Initial	49	27	55.1%	2	4.1%	5	10.2%
3 Month	40	23	57.5%	2	5.0%	1	2.5%
6 Month	36	19	52.8%	0	0%	1	2.8%
9 Month	35	18	51.4%	1	2.9%	1	2.9%
12 Month	34	12	35.3%	0	0%	0	0%

The EQ-5D was more likely to be completed in full than the Capability Questionnaire. It is interesting that as many as 10% of respondents failed to complete the EQ VAS.

Tables 10.4, 10.5 and 10.6 give a more detailed, item by item, breakdown of where there were gaps in responses. In the final column of each table is the percentage of the total number of possible observations which are missing. So, for example, with five items and 49 respondents, there are a possible 245 observations at the initial stage. Six of these observations were missing (2.4%). The capability questionnaire is divided in to the nine sections; each item relates to (is grouped under) one of nine higher-order capabilities.

Table 10.4 Number of Entries Missing: EQ-5D

Stage	Number of Entries Missing by Attribute					%
	Mobility	Self-Care	Usual Activities	Pain	Anxiety & Depression	
Initial	1	1	2	1	1	2.45%
3 Month	1	2	1	1	1	3.00%
6 Month	0	0	0	0	0	0.00%
9 Month	0	1	0	0	0	0.57%
12 Month	0	0	0	0	0	0.00%

Table 10.5 Number of Entries Missing: Initial Capability Questionnaire

Capability / Section	Number of Entries Missing by Question							%
	A	B	C	D	E	F	G	
Self-Respect	0	0	2	0	0	0	0	0.58%
Social Interaction	1	1	1	1	3			2.86%
Parent/ Grandparent	5	5	10					13.61%
Physically & Mentally Active	1	1	1					2.04%
Identity	2	2	2	1				3.57%
Independence & Control	2	2	1	3	3	1	1	3.79%
Relationships	6	6	7					12.93%
Physical & Mental Well-being	1	2	2	1	1	0		2.38%
Enjoyment	4	5	5	3				8.67%

The percentage of questions which were missed within four sections of the Initial Capability Questionnaire are comparable to the percentage of questions missed on the EQ-5D at the initial (zero month) stage. Five of the nine sections (Parent/Grandparent; Identity; Independence & Control; Relationships; Enjoyment) seem to have caused more difficulty for respondents. Three sections stand out as containing a very high number of missing responses: Parent/Grandparent, Relationships, and Enjoyment. The two individual questions which were missed most frequently at the initial stage were:

3. C. "Over the past 3 months, I have been (or would have been) able to support my family financially and provide the financial security that I was able to provide before my pain"

7. C. "Over the past month, I have been able to (would have been able to) enjoy physical affection"

Table 10.6 Number of Entries Missing: Follow-Up Capability Questionnaires

Capability	Number of Entries Missing by Question							Total
	A	B	C	D	E	F	G	
3-Month Follow-up								
Self-Respect	0	0	0	0	1	2	2	1.79%
Social Interaction	0	0	1	0	0			0.50%
Parent/ Grandparent	2	1	6					7.50%
Physically & Mentally Active	0	1	0					0.83%
Identity	1	2	1	1				3.13%
Independence & Control	1	1	0	0	2	0	0	1.43%
Relationships	6	3	7					13.33%
Physical & Mental Well-being	0	1	2	0	0	0		1.25%
Enjoyment	1	1	1	1				2.50%
6-Month follow-up								
Self-Respect	1	1	2	2	1	2	1	3.97%
Social Interaction	2	0	0	0	0			1.11%
Parent/ Grandparent	2	1	6					8.33%
Physically & Mentally Active	1	1	1					2.78%
Identity	1	1	2	1				3.47%
Independence & Control	2	1	1	2	1	1	1	3.57%
Relationships	5	5	4					12.96%
Physical & Mental Well-being	0	0	0	0	1	0		0.46%
Enjoyment	1	2	2	1				4.17%
9-Month Follow-up								
Self-Respect	0	0	2	0	0	0	1	1.22%
Social Interaction	1	1	2	1	1			3.43%
Parent/ Grandparent	4	4	7					14.29%
Physically & Mentally Active	1	1	0					1.90%
Identity	2	1	1	1				3.57%
Independence & Control	0	0	0	0	2	3	3	3.27%
Relationships	3	6	6					14.29%
Physical & Mental Well-being	0	0	0	0	0	0		0%
Enjoyment	0	0	1	1				1.43%
12-Month Follow-up								
Self-Respect	0	0	0	0	0	0	0	0%
Social Interaction	1	1	1	1	0			2.35%
Parent/ Grandparent	1	1	3					4.90%
Physically & Mentally Active	1	1	1					2.94%
Identity	0	0	0	0				0%
Independence & Control	0	0	0	0	0	0	0	0%
Relationships	2	2	3					6.86%
Physical & Mental Well-being	0	0	0	1	0	0		0.49%
Enjoyment	0	0	0	0				0%

At the three-month follow-up stage, there are seven sections of the Capability questionnaire which have a similar percentage of missing responses to those

missed on the EQ-5D (3%). The same two sections as before contain the highest number of questions left blank (Parent/Grandparent and Relationships). In later months the percentage of missing responses on the EQ-5D dips, as does the percentage of missing responses on the Capability Questionnaire, but by a less noticeable extent. The same two sections persistently cause problems. There are some sections, specifically Well-Being, which are, however, largely answered in full.

10.4.2 Exploring Respondents' Understanding of the Questions

Two sections have been identified as having low item response rates, the role of the parent and grandparent and relationships. As well as indicating their responses by selecting an option, some respondents also wrote comments on the questionnaires; these may give a good indication of whether respondents understood the questions and how to answer them. The two sections named above were scanned through on the three-month questionnaires to find additional comments added by respondents. The questions within these sections are listed below, under which are the comments made by respondents.

3. A. Over the past month, I have been able to (or would have been able to) pick up, physically protect and hug young children

“N/A” ID: 019 (54 year old male)

3. B. Over the past month, I have been able to (or would have been able to) take part in games/activities/sports with children

“N/A” ID: 019 (54 year old male)

3. C. Over the past 3 months, I have been (or would have been) able to support my family financially and provide the financial security that I was able to provide before my pain

"n.a" ID: 007 (70 year old male)

"N/A" ID: 008 (62 year old male)

"RETIRED" ID: 017 (87 year old female)

"Have not been able to work for 15 years due to a stroke" ID: 031 (54 year old male)

"N/A – Retired" ID: 040 (70 year old male)

"N/A" ID: 050 (42 year old female)

7. A. I have had the opportunity to meet potential partners/ develop new relationships

"Not interested. Recently widowed" ID: 048 (75 year old female)

7. A. I have felt bad about my partner "missing out" due to my pain

"I am a Widow" ID: 003 (76 year old female)

"Widower" ID: 040 (70 year old male)

"N/A" ID: 046 (64 year old female)

7. B. I have (would have) had as much money to go out and enjoy time with a potential partner/date, as I had before my pain

7. B. My pain has meant that I have shown less patience and emotional support towards my partner than I feel is necessary for a good relationship

"Widower" ID: 040 (70 year old male)

"N/A" ID: 046 (64 year old female)

7. C. Over the past month, I have been able to (would have been able to) enjoy physical affection

"N.A" ID: 017 (87 year old female)

"N/A" ID: 046 (64 year old female)

Ten respondents made comments beside the questions in these two sections of the questionnaire, several of those making similar comments beside more than one question. Several of those who missed questions relating to relationships gave as their reason for doing so the fact they are widowers. Three of those who commented on the question relating to financial support for the family were retired.

10.4.3 Comparability between the two instruments

At this stage, analysis was conducted on the raw data (i.e. entered directly from the questionnaires) without any weighting. Recall that responses on each individual attribute were originally entered as 0, 1, 2 or 3. A zero was entered if the respondent had selected level one (i.e. no problem), a one was entered if the second level was selected (i.e. if there was minor or infrequent difficulty/discomfort), and so on. It is this analysis which will be reported here.

Respondents were allocated to one of four groups according to the EQ-5D tariff value (MVH Group 1995) associated with their health state. Details of the four groups are summarised in Table 10.7. At the initial stage, it was not possible to read off a tariff value for two of the 49 respondents as they had missed at least one question on EQ-5D.

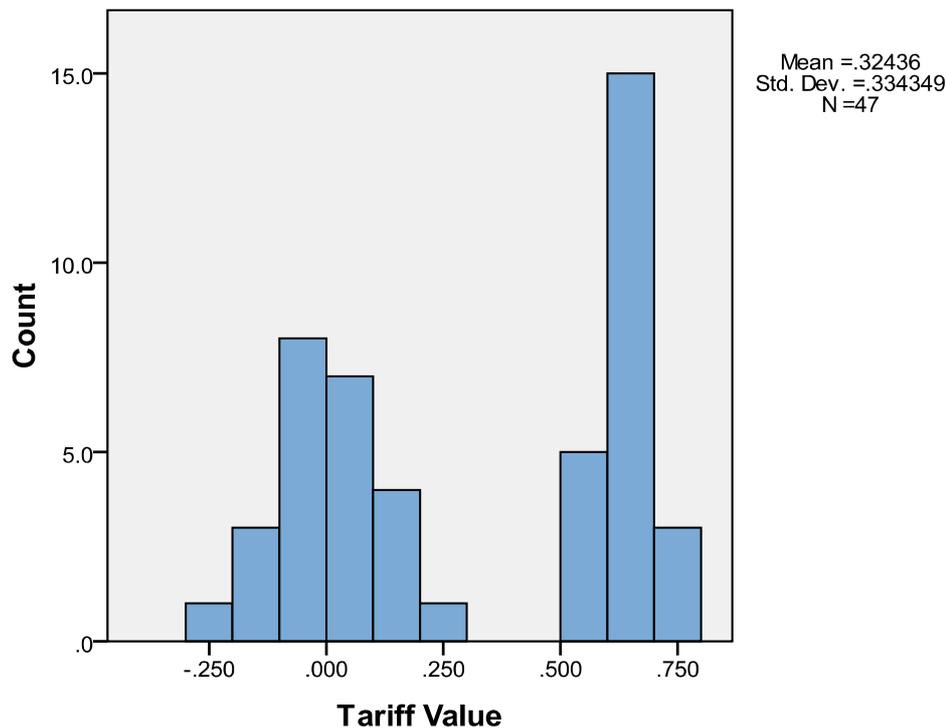
Table 10.7: Summary of the Four Groups Defined by Tariff Values at Initial Stage

Group	Tariff Range	Actual Mean Tariff Value	Actual Median of Tariff Value	Number of Subjects in group	Mean age of subjects	Percentage of male subjects
1	0.538 – 0.796	0.664	0.673	20	59	45%
2	0.279 – 0.537	0.516	0.516	3	47	67%
3	0.020 – 0.279	0.125	0.088	12	59	52%
4	-0.239 – 0.019	-0.090	-0.074	12	57	58%

Figure 10.1 illustrates the overall distribution of EuroQol tariff values. It can be seen that there is a divide between respondents, with approximately half of the respondents having a tariff value above 0.500 and half below 0.250. There is a peak in the distribution at tariff values of around 0.600 (seven

respondents have a tariff value of 0.620). Of the respondents who have the worst quality of life, as indicated by EQ-5D, there is a peak of tariff values below zero (indicating quality of life worse than dead). All 12 respondents in group four have a tariff value of less than zero.

Figure 10.1: Histogram Illustrating the Overall Distribution of EuroQol Tariff Values



Tariff values for the sample population ranged from 0.796 to -0.239. To give some illustration of the range of health states to which these tariff values relate, a tariff value of 0.796 relates to a state in which the subject has:

No problems walking about; no problems with self-care; no problems with performing their usual activities; moderate pain or discomfort; no anxiety or depression.

And a tariff value of -0.239 relates to a health state in which the subject:

Has some problems walking about; has some problems washing or dressing themselves; is unable to perform their usual activities; has extreme pain or discomfort; is extremely anxious or depressed.

Approximately one quarter of the total sample population falls into the fourth group, with the lowest tariff values.

We can now look at responses from the capability-based questionnaire and reach an intuitive conclusion as to whether quality of life appears to be worse for subjects whose health state – as defined by EQ-5D – is associated with a lower tariff value.

The ‘scores’ for all attributes in a particular section were added together and divided by the number of attributes to give a value in the range zero to three. A higher mean value is associated with a greater degree of constraint on a respondent’s capability to enjoy a good quality of life (i.e. a higher value represents a lower ‘capability’). A Summary of the mean scores for the nine sections/capabilities is provided in Table 10.8.

Table 10.8: Responses from the Capability-Based Questionnaire by Group at Initial stage

Section	Total Sample Population		Group 1 (Best Health)		Group 4 (Worst Health)	
	Mean	Median	Mean	Median	Mean	Median
Self-Respect	0.98	0.93	0.79	0.86	1.38	1.29
Social Interaction	1.61	1.80	1.22	1.10	2.25	2.30
Role of Parent/Grandparent	1.91	2.00	1.51	1.67	2.47	2.67
Physically & Mentally Active	2.15	2.00	1.78	1.83	2.53	2.67
Identity	1.33	1.25	0.83	0.75	1.65	1.50
Independence & Control	1.30	1.29	0.84	0.86	1.84	1.86
Relationships	1.59	1.00	1.19	1.00	2.25	2.50
Physical & Mental Well-Being	1.69	1.67	1.45	1.50	2.11	2.13
Enjoyment	1.70	1.75	1.28	1.00	2.21	2.13

We can, based on the data in Table 10.8, say intuitively that there appears to be much agreement between the tariff values (associated with health-states defined by EQ-5D) and the ‘scores’ from the capability-based questionnaire. But, can the capability questionnaire provide us with any more of a complete picture of quality of life for the respondents? Two dimensions from the

capability questionnaire have been arbitrarily selected to give more of a flavour of what it is able to show, summarised in Tables 10.9 and 10.10.

Table 10.9 shows the number of respondents from groups 1 and 4 who selected each level in response to the question relating to being able to do things considered to be worthwhile and productive, taken from the section on Self-Respect. It can be seen that there is quite a clear division between the levels selected by respondents in group 1 and those in group 4, with those with higher tariff values (better health) selecting the levels indicating least problem and those with the worst tariff values selecting those levels indicating more severe problems, as expected. This distinction is not so clear for social interaction, where despite relatively high scores on the EQ-5D respondents in Group 1 score badly in terms of capability.

Table 10.9: Ability to do things considered to be worthwhile & productive

Description	Frequency	
	Group 1	Group 4
Respondent was able to do things they considered to be worthwhile and productive as often as they wanted	6	-
With some restriction and difficulty	12	2
Rarely, and with great difficulty	1	4
Not at all	1	6

Table 10.10: Social Interaction

Description	Frequency	
	Group 1	Group 4
Respondent was able to go out and socialise in the same way and as often as they could before their pain	2	-
Most of what they could before and as often	4	-
Little of what they could before and less frequently	11	3
Respondent tended to stay at home due to their pain	2	9

10.5 Discussion

It is recognised that the Capability Questionnaire is a lengthy questionnaire, and using a questionnaire with forty-two questions (see Appendix C) increases the risk that respondents will become fatigued and/or simply bored. It also increases the likelihood that careless mistakes are made such as turning over two pages at once, or missing individual questions or sections

due to other distractions. It is also recognised that the move beyond health in order to measure wider quality of life means that there may be questions which do not seem immediately relevant to all respondents, although the inclusion of all questions was carefully considered during the qualitative work.

The sections which caused the greatest difficulty for respondents, or which some respondents didn't appear to identify and engage with, were the role of parent and grandparent and relationships. Concerns expressed in the focus groups which related to parenting included the fact that pain disrupted what would ordinarily be any regular flow of income coming into the household. Periods of earning low or no income was reported to lead to difficulties planning for the future (e.g. building up funding for children to go to university), generally making it difficult to provide financial stability within the core family group (extending to grandchildren), and even with simple things such as buying birthday presents. Other concerns were not being able to join in activities with children (cycling, playing football, etc), which in turn led to feelings by the parent/grandparent that they were not perceived to be 'fun', that they were not approached to help with or to be involved in other activities/issues that the parent/grandparent felt they *did* have the capability to do/help with.

It is clear that chronic pain has a significant impact on how sufferers cope with and fulfil the role of parent or grandparent and on the relationship with the child. The difficulty is also clear, in that not every respondent will have this role and this relationship, or indeed ever want it. The questions are phrased in terms of capability rather than functioning, but it may simply be a step too far to expect respondents to imagine their ability to perform a role that they genuinely have no experience of. Similarly respondents were asked how their pain will impact on a bond/relationship which, in perhaps a significant number of cases, they are simply not interested in forming.

One possible solution to the problem may be to miss this section out completely. This first solution would result in an important aspect of life for many respondents being excluded from the questionnaire. Another option

would be to try and phrase the question in a way which is broader than it is currently phrased. Respondents could perhaps be asked about the impact their pain has on their ability to support the upbringing and the future well-being of children within their family. This second solution while representing an effort to make the question relevant to a wider range of respondents may still, however, be missed by respondents who do not have children of their own.

The Capability 'Relationships' caused a great deal of difficulty during the design of the questionnaire as respondents have very different experiences depending on if they are single or in a relationship. As a section of the questionnaire it has also been frequently missed. This may be because it is not obvious how someone who is widowed, or in the case of one respondent who for religious reasons had chosen to be celibate, should answer the question. Grewal, Coast *et al.* (Grewal, Lewis *et al.* 2006; Coast, Peters *et al.* 2008) chose to have a single capability 'Attachment', relating to love *and* friendship and this would seem a simple solution. However, when patients in the focus groups discussed relationships they discussed a small number of quite specific issues which seem different from the issues likely to be raised by older people.

The patients reported how their pain impacted also on the life of their partners and of the guilt and upset that they experienced because of this. It is because of the intensity of the relationship, the time spent together and the greater degree of emotional dependency, that their pain was able to have such impact on the life of another. Some respondents who were not in a relationship spoke of how, because their pain limited their ability to work and socialise, their chances of meeting potential partners was limited compared to that of a healthier person. Some also reported fearing that they would be viewed as someone who would have to be 'looked after'. While the dimension "I can have all of the love and friendship that I want" would appear to appropriately summarise what is important to those who are single and would be more likely to be answered by those who are widowed, it doesn't appear to encompass the concerns of the pain sufferers from this project who were in a relationship.

Issues relating to friendship appeared, for them, to be distinctly separate from concerns relating to relationships.

As for the section on the role of the parent or grandparent, there are two obvious solutions: (i) either cut out this section, which focuses on what will be for many respondents an important aspect of their life; (ii) or attempt to make the section more relevant to all respondents.

The questions developed through the qualitative work are concentrated on each specific issue identified from the focus groups, and this is the most likely reason for the high number of questions that were missed during the piloting stage. The questions on relationships are a good example of this, even to the extent that different questions were asked to those who were single as opposed to those in a relationship. On reflection, the entire section on Relationships could have been effectively replaced by two questions, the first along the lines of that used by Grewal, Coast *et al.* (“I can have all of the love and friendship that I want”), the second a question such as: “My health has changed the nature of the relationship that I have with those closest to me”. The specific nature of many of the questions, and because of this the need to include such a high volume of questions, is likely to be an indication of weakness in the analysis of the qualitative data, at the stage of identifying and describing the main themes. If the qualitative work was revisited then the number of dimensions (questions) could probably be reduced, and perhaps even the number of higher-order capabilities.

It is felt that although the level of detail or questioning within the entire questionnaire may have been too specific, and although the inclusion of certain Capabilities needs to be reviewed, that no changes should be made without first revisiting the qualitative data. It would be wrong to exclude functionings and capabilities which are important to sufferers. The option of cutting out entire sections due to poor response rates must therefore be rejected. Instead the process through which the capabilities were identified and refined should be reviewed.

It should be remembered that sections such as those relating to Self-Respect, Physically and Mentally Active, and Physical and Mental Wellbeing, were answered fully by almost all respondents, and certainly by a comparable percentage to those who completed the EQ-5D fully.

On the whole, results which were obtained from the Capability based questionnaire appear useful and appear to provide an accurate summary of changes in quality of life for individuals and for the group over the twelve month period. Results from the EQ-5D generally support those from the Capability questionnaire, although the Capability questionnaire can offer a fuller picture of quality of life.

Chapter 11: Study Phase III: Weighting Capabilities

In chapter seven a number of possible methods were discussed, from within both a Capability and a health economics context, in an attempt to arrive at a suitable and acceptable method for deriving a scoring system for the Capability-based Questionnaire. It was decided that the best option in general would be to obtain relative weights for the capabilities and then to combine these with values for the levels within the functionings. The chosen method will be heavily based on a Multi-Attribute Value (MAV) Method originally derived and used by Peacock et al (2007).

Section 11.1 gives an overview of the study, sections 11.2 and 11.3 discuss sample size and recruitment. It was noted in chapter seven that the MAV method used by Peacock *et al.* involves six steps, the first two of which are similar to those undertaken so far in this study. Step three involves a panel scaling the different levels within the attributes, and is detailed in section 11.4. For step four, the panel then assesses the relative importance of each attribute, through the use of a 'swing weights' method. Details of how step four has been implemented are given in section 11.5. Section 11.6 addresses a number of study design issues which it would be important to consider should the method be applied on a larger scale.

Step five involves evaluating how programmes impact on respondents in terms of the nine higher-order capabilities. Chapter 10 illustrated how responses to the questionnaire indicate respondents' achievement in terms of the nine higher-order capabilities. Step six involves estimating the capability weights and choosing a functional form for the MAV model, which will be discussed in the latter half of this chapter. Part-whole bias appears to have been present in the results and so additional work was conducted to further explore this phenomenon. Section 11.7 therefore reports an extension to the original study design for step four and section 11.8 discusses how changing the study design affected the results, in terms of minimising part-whole bias. Section 11.9 summarises.

11.1 Study Overview

In this study, the 'panel' was members of the general public who completed stages three and four individually, in the form of two paper-based exercises, but who were taken through the prior explanation of the exercises and were involved in the ensuing discussion within small groups.

Groups lasted for approximately 90 minutes and were run in four distinct parts. First, participants were welcomed and given the background to the study; they were told of how the questionnaire had been developed and were given a definition of chronic pain (as used throughout this thesis). Next, participants were given the scaling exercise (stage three), which was explained to the group as a whole, completed by participants individually and then discussed. After this, participants were given the swing-weighting exercise (stage four), which again was explained, completed by the participants individually and then discussed.

An important point to note when it comes to implementing the third step of the MAV method is that with the Capability questionnaire there are dimensions (functionings) grouped together under categories (higher-order capabilities), effectively creating a 'middle' tier which is additional to the two tiers used by Peacock and Richardson (which were, levels within attributes and overall attributes). Here, step three will involve scaling levels on the functionings, and step four will involve quantifying trade-offs between capabilities. It will be assumed that each of the functionings that make up a particular capability are equally important/valuable.

For step three each functioning was presented to respondents as an individual question. For each question there was a scale of zero to 100, with the worst level being fixed at zero on the scale and the best level being fixed at 100. Participants were asked to place the two remaining (intermediate) levels of the functioning somewhere on the scale.

Step four involved participants being presented with four scenarios. The first was that in which all nine capabilities were at best (for top-down), or all at worst (for bottom-up); half of the groups were given top-down, the other half bottom-up. The starting scenario was already fixed on the scale (at either zero or 100). In the following three scenarios up to three of the nine capabilities varied and participants considered these scenarios and placed each of them at a position on the scale. The scaling exercise and the swing-weighting exercise will be described in more detail below.

11.2 Sample Size

The task was presented to small groups of participants. The advantages of inviting participants along to make up small groups were that the task could be explained to numerous participants at once, so making efficient use of time, and also that the tasks could be discussed within the group after completion.

As stated above, step three involves a panel scaling the different levels within the functionings. There are 42 functionings within the questionnaire, each with four levels. Each of the levels within each of the 42 functionings had to be valued directly as each is unique. The scaling method is relatively simple and so it was felt that, providing the functionings were divided over a number of groups, it would be reasonable to expect to obtain values for all 42, even in a small pilot study. It was decided to present respondents with seven dimensions each, which meant that a minimum of six groups was required.

While respondents worked alone for most of the time, there was a discussion after each of the two exercises and so it was felt that the groups should be small enough to allow for free discussion, to which respondents felt comfortable contributing. Also, it was expected that respondents would ask the researcher for guidance during the scaling exercise and swing-weighting exercise and so it was agreed that the groups should be kept to a manageable size, so that assistance could effectively be offered to everybody. It was therefore decided that five to six participants would be recruited for

each group, with each scenario being considered by one group. With six groups, this meant a target sample size of up to 36 participants.

In the study by Peacock *et al.* (2007), the only study in which, to our knowledge, this method has been used, the panel consisted of 19 experts, and panel members appear to have completed both the top-down and bottom-up version of the task. A relatively small sample size would be expected, given that the researchers were limited by the number of experts working in the field, and presumably by the work pressures of those experts. What is more, the researchers only had three attributes to work with. In our study the panel was made up of members of the general population, and there are a greater number of attributes (capabilities), and so it was felt that a larger sample size was appropriate.

11.3 Recruitment

The study was reviewed and approved by the Faculty of Health Ethics Committee at the University of East Anglia. On the instruction of the Ethics Committee, Invitation Letters, Information Sheets and Consent Forms (see Appendix D), together with pre-paid envelopes, were sent out in the post (from 14th April 2008 onwards); only when a Consent Form was returned by a potential participant was the individual contacted by telephone and a date agreed on which the individual would attend a group.

Initially, all participants were recruited from a database of research volunteers maintained by the School of Economics at the University of East Anglia. The database contained the details of members of the public living in the Norwich area who had previously indicated that they would be willing to participate in academic research; many had taken part in previous studies conducted by the University of East Anglia. Due to the nature of recruitment (using the postal service) the number of willing participants returning Consent Forms was relatively low; because of this a small number of undergraduate and postgraduate students were also recruited for the study. Participants were told in advance that they would receive £15 for taking part.

Seven groups were held in total, between 6th May and 23rd May 2008. There were between two and five people in each group. Due to poor attendance in Group four, an additional group was held, effectively to repeat this group with a larger number of participants (and hence seven groups were held in total), these two groups are reported as if they were a single group, Group four.

Twenty-seven participants took part in the study. Characteristics of the participants are reported by group in Table 11.1 and summarised across the entire sample in Table 11.2. TD denotes that a group were presented with the top-down approach; BU denotes that the group were presented with the bottom-up approach.

Table 11.1 Summary of Characteristics of Participants by Group:

Group	No. Male Participants	No. Female Participants	Min Age	Max Age	Mean Age
1 (TD)	3	2	49	65	56
2 (BU)	0	2	30	66	48
3 (TD)	3	2	34	67	57
4 (BU)	2	4	25	64	45
5 (TD)	3	2	28	63	50
6 (BU)	0	4	23	41	29

Table 11.2 Summary of Participant Characteristics:

Characteristic	Frequency	Percentage	Percentage for Norwich Population (From 2001 UK Census)*
Male	11	40.74%	47.81%
Female	16	59.26%	52.19%
Aged 20 - 35	8	29.63%	34.17%
Aged 36 - 50	5	18.52%	24.65%
Aged 51 +	14	51.85%	41.18%
Employed	8	29.63%	-
Retired	10	37.04%	-
Student	6	22.22%	-
Other	3	11.11%	-

* For Population aged 20 years and over

The sample used was a convenience sample, although some effort was made, where practical, to ensure that a range of segments of the local population were represented in the groups. It can be seen that, in comparison

to the local population as a whole, males were slightly underrepresented in the sample, and those aged over 51 years were slightly overrepresented, mostly at the cost of those aged 36 to 50.

11.4 Scaling Exercise for Levels

Participants were given a pack, which contained a brief amount of text about the background to the study and the purpose of the scaling exercise, an example question and an example scale, all on coloured paper. This was explained by the researcher in the same order as it was presented in the Participant Pack (see Appendix D), and PowerPoint was also used to assist with the presentation. For five groups the example question related to self-care (see Figure 11.1). The group who had self-care within their 'real' questions were given a different example question.

Figure 11.1: Example Question

Example Question

Over the past month, I have been able to care for myself (dress, shower, and use the toilet):

- With no help at all from others and no difficulty
- With no help at all, but with some difficulty
- I have needed some help and/or have great difficulty dressing, showering or using the toilet
- I am completely dependent on others to dress, shower, or use the toilet

It was explained that the top level on all of the questions is always the best possible level, and as such will be fixed at the top of the scale, at 100. It was also explained that the bottom level on every question is the worst possible level and so will be fixed at the very bottom of the scale, at zero. It was then explained that the two middle levels fitted at points somewhere on the scale between these two extremes, and that the position of the two middle points would be determined by the participants themselves.

The researcher then presented a scenario unrelated to chronic pain in order to familiarise participants with the rating scale. The example related to the quality of a dinner available to the participant. Four levels, or qualities, of dinner were stipulated: (i) dinner in the participant's favourite restaurant, (ii) a simple pub lunch, (iii) a home-cooked meal, (iv) a pre-packed sandwich. A number of points were stressed as the example was discussed:

- That the respondent should answer in a way that reflects their *own personal* preference (and that one person's tastes will often differ from those of another).
- That the respondent should answer as if *they* were the one's facing the scenarios described (i.e. as if it was *you* that was eating the pub lunch or the sandwich).
- That it did not matter in what order the middle two levels were placed (e.g. pub lunch above home cooked meal, or vice-versa).
- That they should not try to introduce related factors such as what the cost may be, or the opportunity cost associated with their own time.

Once respondents were clear about the concept of the rating scale they were asked to work through seven questions from the 42 contained on the questionnaire. In their packs, on white paper they were given seven questions to work through themselves, individually. They were encouraged to ask questions if any clarification was needed. For each question they were given a description of the two middle (or 'missing') levels (which were labelled A and B) and a rating scale of zero to 100. On each rating scale the best level was pre-printed on at 100 and the worst level was pre-printed on at zero. Every participant in a group was given the same seven questions; each group was given a different set of seven questions, although one set was given to two groups due to poor attendance in the first (hence seven groups rather than six). As with the 'lunch example', respondents had the option of reversing the order of the two middle levels if they wanted to (i.e. placing the third level higher on the scale than the second).

When every participant had completed the task they were encouraged to discuss their answers, and a series of open-ended questions were used to initiate this discussion. As this was the first time that the questions had been presented to the general public, the sample group were unlikely to have chronic pain and hence there was a need to establish whether respondents had understood the questions. As this was a pilot study, there was also a need to establish if respondents had understood how to complete the task itself. It was further felt that discussion with respondents may go some way to revealing why they had chosen the answers that they had. For example, if there was something in the wording of a question that had influenced their answer, if they were drawing on “external considerations” (such as experience of caring for a dependent relative), or if they had simply fallen into a repetitive pattern when considering all seven questions.

Loosely, the questions used to initiate discussion were:

1. How did people find the task? Were the instructions clear?
2. Were there any descriptions within the questions that were confusing?
3. Were there any questions that you had to think about for a long time; and, if so, why was this?
4. If we take Question ‘X’ as an example, would anybody like to volunteer where they put levels A and B on the scale?
5. Why did people choose the points that they did?
6. Has anybody got any other comments they would like to offer?

An audio recording was made of the discussion so that the researcher could compare and detect any similarities or disagreements arising within the discussion between different group members, or between different groups.

11.5 Swing-Weighting Exercise

The purpose of the swing-weighting exercise was to quantify trade-offs between the nine capabilities. Participants were given a second Pack for the swing-weighting exercise (Appendix D). Once again, the first (coloured) pages of the pack explained the background to and the purpose of the task

and contained an example, and, once again, the researcher talked through the task with the aid of PowerPoint. It was explained that each of the questions on the questionnaire was grouped within one of nine categories. It was further explained that when a respondent selects the top level on every question within a particular category then that category can be said to be 'at best', and that, vice-versa, when a respondent selects the bottom-level for every question within a particular category then that category can be said to be 'at worst'. Participants were told that they would be asked to focus on scenarios in which each category was at one of these extremes.

It was then explained that when all nine categories are 'at best' a respondent can be said to have the best possible quality of life and, vice-versa, when all nine categories are 'at worst' the respondent should be thought of as having the worst possible quality of life. The researcher stressed that the scenario in which all categories are 'at best' is fixed at 100 on the scale, and that the scenario in which all categories are 'at worst' is fixed at zero on the scale. Participants were then told that they would be asked to consider four scenarios, the first was all at best or all at worst (depending on whether the group was doing top-down or bottom-up), and would be fixed on the scale as just indicated; the other three scenarios would involve some categories at best while others were at worst and Participants would be asked to decide where on the scale these scenarios should be placed.

To familiarise participants with the scale in this context the dinner example was extended. Participants were reminded that in the last example they were asked to focus only on one aspect of a normal day, the quality of dinner available to them. In this example the dinner would combine with other factors (e.g. the weather, and the quality of time spent with family) to determine the quality of the overall day. If they had enjoyed the best possible dinner, the weather had been as good as possible and the quality of time spent with family had been high then the day as a whole would be as good as it possibly could be, and vice-versa. They were then encouraged to think of a day in which the weather was awful, they had had an argument with their family but then enjoyed dinner in their favourite restaurant. They were

encouraged to think about where the day described here would fit on the scale of zero to 100, considering to what extent the dinner would compensate for the rest of the day being bad, and to what extent the awful weather and the argument would make the whole day 'bad'.

Once participants understood the scale they were read a description of the nine categories in their opening state (i.e. all at best for top-down, or all at-worst for bottom-up) and encouraged to seek clarification if the descriptions were at all unclear. They were then asked to complete the task individually. For the task they were given a full summary of the starting Scenario (labelled Scenario A), and were then given a summary of any categories that had changed for Scenarios B, C and D. On the last page was a scale, on which Scenario A was pre-printed at zero or 100. The pages for the actual task were printed in colour on plain white paper. Categories at best were printed with a heading in green text, those at worst were printed with a heading in red text.

A short discussion was then initiated following a similar format to the discussion held after the scaling exercise. Once again, an audio recording was made of the discussion. After the discussion participants were given the option of giving different responses on a new but identical scale.

There has been much emphasis placed within the Capability literature on arriving at lists of capabilities and deciding which capabilities are of importance (and how important) by open discussion and deliberation. In this study the list of capabilities was derived through direct discussion with chronic pain patients. Exactly what would constitute open discussion and deliberation when it comes to developing a scoring system hasn't been defined. Here it is interpreted to be sufficient to take the average of a number of individual responses. It was felt that it would be of interest to see how engaging in an open discussion would influence people's responses. It could be assumed that discussing their responses as a group would encourage respondents to consider their approach to the task in new ways and to consider factors that had not occurred to them previously. It would then be expected that, following

discussion, responses would, to some extent, converge and that preferences would be more likely to pass tests of dominance, consistency, etc. If responses remain very different then this is an indication that agreement and convergence on a joint response within a group of people is unlikely ever to be achieved.

A problem was encountered with the three capabilities chosen to vary. When Capabilities were swung to the extremes of 'at best' or 'at worst' participants found that there was some contradiction in terms of what an individual was described as being able to do in some categories and what they were unable to do in others. For example, with Independence & Control at worst but other categories at best an individual would need help to dress and shower, but then be described as not having any need to ask for special treatment, not being treated any differently to other people and being able to get as much physical exercise as they wish. Although participants were still able to complete the task in a way that they felt was meaningful, the scenarios were changed after the first top-down and the first bottom-up group in an attempt to make them more plausible. So, for example, with top-down, up to three Capabilities were chosen to swing to at worst from a starting point of all at best. An effort was made to select three capabilities which didn't contain functionings which, when at worst, would seemingly contradict with the fact that other functionings (related to other capabilities) were simultaneously at best. This was just an effort to make scenarios more plausible, so that they didn't describe a situation where, for example, a person was unable to dress them-selves but could take part in physically demanding leisure activities.

It has been noted previously that if the nine higher-order capabilities are mutually independent then an additive functional form will be appropriate for the model. Discussion from the groups will be an important indication of whether respondents felt that the higher-order capabilities are indeed mutually independent, although as discussed by Keeney and Raiffa (1993) an additive functional form may still be an acceptable approximation.

11.6 Study Design & Statistical Methods

In terms of analysis for stage three, nothing more complicated is required than to estimate the mean and median values across each of the dimension levels, together with the range.

In terms of stage four, the fact that there are nine capabilities which could each be 'at best' or 'at worst' means that there are a total of $2^9 = 512$ possible scenarios for the swing weighting. The 512 scenarios would, however, include the scenario in which all nine capabilities are at best and that in which all nine capabilities are at worst, which it has already been determined will take the values 100 and zero respectively. It is unlikely given constraints such as time and funding that it would be feasible to directly value all 510 scenarios directly, in any study, and so a decision must be made as to the lowest number of scenarios which could be valued and still ensure good study design with workable analysis. This number will largely be determined by the nature of the analysis that is planned.

Peacock and Richardson (2007) use a relatively straightforward method for calculating weights for their attributes, summarised in three steps:

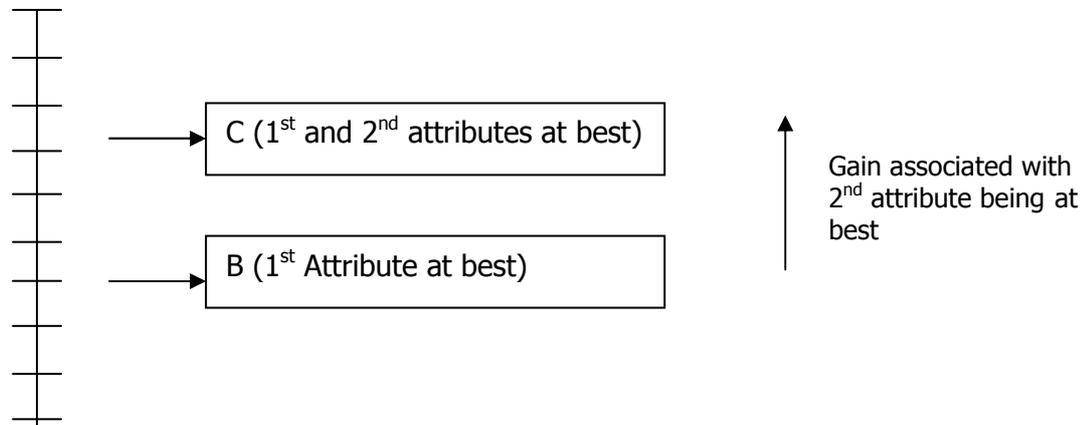
1. Where a scenario represents an improvement (deterioration) in one attribute independently, then the mean gain (loss) associated with that one scenario is taken as the mean utility gain (loss) associated with that particular attribute.
2. Where there is a two-step process (i.e. where scenario B represents a change in the first attribute independently and scenario C represents a change in the first attribute and a second together) then the mean gain (loss) associated with the second attribute is found by calculating the difference between the second shift and the first. For bottom-up, the gain associated with the second attribute improving to at best would be calculated using the formula:

$$\text{Mean Gain} = \text{Score for C} - \text{Score for B (see figure 11.2).}$$

3. The attribute weight is simply the mean of the mean gain from bottom-up and the mean loss from top-down.

Figure 11.2: Calculation of gain in value associated with an improvement in the second attribute

All at Best (100)



All at worst (zero)

As the statistical analysis is no more complex than subtraction and averaging it would seem unnecessary to determine the minimum number of scenarios needed by any technique other than straightforward pragmatism.

Indeed, if the same methods are used as were developed by Peacock and Richardson then only scenarios in which one capability changes individually and in which two capabilities change together need to be considered.

As an example, one could set scenarios in which capabilities 1, 3, 5, and 7 varied independently, along with scenarios in which the capabilities 1 and 2, 3 and 4, 5 and 6, 7 and 8, and 1 and 9 varied jointly. This would mean using nine scenarios for top-down and nine for bottom-up, 18 in total. From just 18 scenarios weights could be calculated for all nine capabilities using the same methods as Peacock and Richardson.

Although such an approach to the study design and statistical analysis would have the advantage of being simple, this could also be a source of weakness.

An alternative means of arriving at weights for the capabilities (and one explored in Section 11.9) would be to use regression analysis. An attempt will

also be made, therefore to estimate weights for the attributes using an analysis of variance (ANOVA) model with no intercept (regression through origin). If its score on the scale is simply used as the raw and uncalibrated weight for a given scenario, then a regression model with score as the dependent variable can be used, with nine dummy explanatory variables representing each capability at worst (0) or at best (1). The coefficients on each of these explanatory variables can then be interpreted as its raw uncalibrated weight.

For the swing-weighting, the minimum number of scenarios required to estimate main effects would be 12. This would mean an orthogonal array of strength two, with 12 runs (scenarios) and nine factors (attributes). However, as well as having 12 scenarios for the form of swing weighting using the top-down method there would also need to be 12 scenarios for the form of swing weighting using the bottom-up method. This means that 24 scenarios would actually be needed, and that there would effectively be a fold-over design.

A 'main effect' is defined as "the difference in the means of each level of a particular attribute and the overall mean, such that the differences sum to zero" (Louviere, Hensher et al. 2000).

Although it was noted earlier that it would be possible to limit the scenarios used to only those in which one capability changes independently or up to two change together, there may be good reasons for using scenarios in which more capabilities change at any one time. For example, it may make it easier to detect trends as the number of capabilities changing in any one scenario increases (such as diminishing effects, or interaction between the capabilities). Furthermore, this may help to make scenarios more plausible for respondents and make the scenarios more varied. As a pilot study it would be useful to be able to explore the results as much as possible and not narrow our focus too much. If a larger study was being undertaken then 12 (or rather 24) scenarios would be the *minimum* which could be considered useful. Best practice stated within the DCE literature is to work with a full factorial design wherever possible and Lancsar and Louviere (2008) warn

against relying on a fractional factorial (a minimum number of scenarios necessary to estimate main effects). Although the two approaches are very different and the MAV method has not been used enough times for this type of best practice to emerge, it would seem appropriate to work with the maximum number of scenarios that time and resources will allow.

Table 11.3: Possible scenarios if the level of only three Capabilities is varied

	SR	PG	ID	RE	IC	EN	SI	PA	WB
<i>TD</i>	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	0	1	1
2	1	1	1	1	1	1	1	1	0
3	1	1	1	1	1	1	1	0	1
4	1	1	1	1	1	1	0	0	0
5	1	1	1	1	1	1	1	0	0
6	1	1	1	1	1	1	0	0	1
7	1	1	1	1	1	1	0	1	0
<i>BU</i>	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	0	1	1	1
12	0	0	0	0	0	0	0	1	1
13	0	0	0	0	0	0	1	1	0
14	0	0	0	0	0	0	1	0	1

SR = Self-respect
 PG = Role of parent/grandparent
 ID = Identity
 RE = Relationships
 WB = Physical & Mental Well-being
 EN = Enjoyment
 SI = Social Interaction
 IC = Independence & Control
 PA = Physically & Mentally Active

1 = 'at best' 0 = 'at worst'

There is a trade-off however, as the more things that change in a particular scenario the more complex the task may become for respondents. To reduce the amount of information that respondents were asked to consider in the pilot study, each respondent was asked to consider scenarios in which the level of up to just three of the nine capabilities was varied. Focussing on only three capabilities meant that the total number of possible scenarios for the pilot study was reduced to $(3^2) - 2 = 7$ (all at best and all at worst are excluded). Because both a top-down and a bottom-up method were used, the number of possible scenarios was actually 14. Some respondents would be asked to consider variation in the level of three capabilities following a starting scenario in which all of the capabilities were 'at best', i.e. using the top-down approach (row TD in Table 11.3), while others would be asked to consider variation in the level of three capabilities following a starting scenario in which all were 'at worst', i.e. using the bottom-up approach (row BU in Table 11.3). It can be

seen that the bottom-up method is effectively a fold-over of the top-down method.

To derive a complete index of capability within a larger study it would not be enough to only vary the same three capabilities and a *minimum* of 12 scenarios would need to be selected which reflect the wider range of possible combinations. One possible method of selecting 12 scenarios from the 510 possible scenarios while still ensuring that in each scenario a maximum of three capabilities are allowed to vary would be to:

1. make four selections of three capabilities (three from the nine). These will be the capabilities allowed to swing between best and worst.
2. label the selections A to D
3. for each of the selections (A to D) pick three of the seven possible scenarios
4. these scenarios can be labelled A1 to A3, B1 to B3, C1 to C3 and D1 to D3

For example, our selection for A could be: Social Interaction, Physical & Mental Well-being and Physically & Mentally Active (as above). If only these three capabilities are allowed to swing between best and worst and the remaining six are 'fixed' at best then we have the seven (one to seven) possible scenarios listed in Table 10.3. From these seven we would choose three, for example scenarios 1, 4 and 5. These could be re-labelled A1, A2 and A3, and were chosen as the scenarios for the pilot study, see Table 11.4.

Table 11.4: Scenarios for Pilot of Swing-Weighting

	SR	PG	ID	RE	IC	EN	SI	PA	WB
A1	1	1	1	1	1	1	0	1	1
A2	1	1	1	1	0	1	0	0	1
A3	1	1	1	1	0	1	1	0	1
FA1	0	0	0	0	0	0	1	0	0
FA2	0	0	0	0	1	0	1	1	0
FA3	0	0	0	0	1	0	0	1	0

In addition, the fold-over of A1 to A3 would also be used, which would be scenarios 8, 11 and 12 from Tables 11.1. We can re-label scenarios 8, 11 and 12 as FA1, FA2, and FA3, see Table 11.4.

Respondents in the pilot study will need to consider the 'starting scenario', i.e. all capabilities at best for top-down and all at worst for bottom-up. The starting scenario will be presented as a list of short summaries of the nine capabilities. Respondents will then consider three further scenarios, in the case of this pilot study either A1 to A3 or FA1 to FA3.

11.7 Results: Scaling Exercise

This section presents data from the scaling exercise for the 42 questions on the questionnaire. Recall that the top level on each question was fixed on the scale at 100, the bottom level fixed at zero. Participants were presented with a description of the two middle levels (labelled A and B) and asked to place these at appropriate positions on the scale. Scores are presented in Table 11.5. One of the four options available for patients to select on the two questions relating to employment is essentially an 'opt out' for those over 60 and retired. Retirement was given as level B to respondents completing the valuation exercise. However, respondents didn't feel able to value retirement relative to either being employed or to experiencing difficulties with employment due to ill health. For this reason level B on these two dimensions (1G and 2A) has been excluded.

Table 11.5: Scores for levels within Dimensions:

Dimension	Level A				Level B			
	Minimum	Maximum	Mean	Median	Minimum	Maximum	Mean	Median
1A Respected by others	80	90	85.0	85.0	10.0	20.0	15.0	15.0
1B Honest about severity of pain	70	90	78.0	70.0	10.0	40.0	24.0	20.0
1C Believed	45	80	66.3	70.0	20.0	40.0	27.5	25.0
1D Family have accepted illness	60	90	80.0	90.0	10.0	70.0	34.0	30.0
1E Help with everyday tasks	60	100	84.0	90.0	10.0	70.0	40.0	50.0
1F Able to do things worthwhile	70	80	75.0	75.0	20.0	20.0	20.0	20.0
1G Able to earn a wage	50	80	65.0	65.0	-	-	-	-
2A Meet people through work	40	70	57.5	60.0	-	-	-	-
2B Go out and socialise	70	80	75.0	75.0	20.0	20.0	20.0	20.0
2C Visit friends & family	60	80	70.6	73.0	10.0	35.0	25.0	30.0
2D Included as part of a social group	40	90	70.0	80.0	00.0	40.0	22.0	20.0
2E Money to go out & socialise	70	80	72.5	70.0	20.0	35.0	26.3	25.0
3A Hug/protect young children	65	90	76.3	75.0	20.0	50.0	32.5	30.0
3B Games/activities with children	40	80	65.0	70.0	10.0	55.0	24.2	20.0
3C Support family financially	75	90	85.0	90.0	18.0	20.0	19.6	20.0
4A Hobbies & Interests	60	70	65.0	65.0	35.0	40.0	37.5	37.5
4B Physically demanding leisure	60	90	76.0	80.0	40.0	60.0	50.0	50.0
4C Paid work	50	70	63.6	70.0	10.0	30.0	24.1	20.0
5A Ask for special treatment	70	90	78.7	76.0	10.0	30.0	20.8	20.0
5B Treated differently	60	90	75.0	75.0	05.0	50.0	22.6	18.0
5C Conversation	70	90	78.0	80.0	10.0	60.0	34.0	40.0
5D Able to disguise/hide pain	60	80	67.5	65.0	20.0	30	27.5	30.0

Table 11.5 Continued

	Minimum	Maximum	Mean	Median	Minimum	Maximum	Mean	Median
6A Self-care	70	80	75.0	75.0	20.0	20.0	20.0	20.0
6B Drive/ use public transport	40	80	62.5	65.0	10.0	65.0	24.2	20.0
6C Walk	50	90	70.0	70.0	20.0	60.0	40.0	50.0
6D Basing routine around medication	40	90	70.8	75.0	10.0	30.0	17.8	19.0
6E Stairs	30	70	56.0	70.0	10.0	50.0	26.0	30.0
6F Ability to plan things in advance	50	80	68.0	70.0	20.0	40.0	30.0	30.0
6G Cook & look after home	40	90	74.0	80.0	20.0	60.0	40.0	40.0
7A(ii) Partner missing out	50	90	80.0	90.0	10.0	30.0	19.0	20.0
7B (ii) Patience & emotional support	50	80	65.8	70.0	10.0	30.0	17.2	15.0
7C Physical Affection	48	80	65.6	60.0	05.0	10.0	09.0	10.0
8A Pain	50	90	75.6	80.0	05.0	23.0	13.6	10.0
8B Effects of medication on future health	60	80	70.0	70.0	10.0	40.0	26.0	30.0
8C Side-effects of medication	70	80	75.0	75.0	30.0	30.0	30.0	30.0
8D Sleep	70	80	73.3	70.0	20.0	35.0	29.2	30.0
8E Physical exercise	60	90	76.0	80.0	20.0	40.0	34.0	40.0
8F Depression	60	80	67.8	68.5	20.0	30.0	23.3	20.0
9A Tiredness & Frustration	80	90	85.0	85.0	10.0	30.0	20.0	20.0
9B Embarrassment	40	70	55.0	55.0	10.0	30.0	20.0	20.0
9C Enjoyment	80	90	86.0	90.0	30.0	80.0	51.0	50.0
9D Concerns about the future	60	90	72.5	70.0	20.0	30.0	27.5	30.0

Once again Question 6A will be considered as an illustration. The top level, being able to care for one's self with no help at all from others and no difficulty, is fixed at 100. The second level (level A), being able to care for one's self with no help at all, but with some difficulty, was given a mean score of 75. The third level (level B), needing some help and/or having great difficulty caring for one's self, was given a mean score of 20. The bottom level, being completely dependent on others to dress, shower, or use the toilet, is fixed at zero. Responses for A and B on this particular dimension fell within a very narrow range indeed (between 70 and 80 for A, and for B all responses were at 20).

Table 11.6: Summary Statistics for Table 10.5

	Minimum Value	Maximum Value	Mean of Minimum Values	Mean of Maximum Values
Level A	30	100	55.0	86.0
Level B	0	80	13.6	51.0

Table 11.6 shows the range of scores across all 42 dimensions. On average, scores for level A tended to fall between 55 and 86 on the scale, although some fell as low as 30 and went up as high as 100. The fact that one respondent gave a score of 100 for level A may indicate that they experienced some confusion. What is noticeable from Table 11.3 is that in many cases (28 out of 42) the score for level A is higher than 70, and this score above 70 for A is frequently accompanied by a score for B which is equal to or less than 30; hence what we tend to see is some degree of polarization between A and B. A likely cause of this is the descriptions of the levels, as the description of A tends to be quite positive ("I can do X with some difficulty") whereas the descriptions of B tend to be quite negative ("I can do *little* of X", or "I *cannot* do X at all some days). Certainly this explanation fits with what participants were saying in the discussion. Participants suggested that, although there may be 'some' difficulty doing something of importance, it was positive that the outcome was still achievable while at the same time the individual maintained their independence, in this way the restriction tended to be viewed as an inconvenience or frustration. Participants tended to view a move to level B as

more of a loss or as a much more significant restriction in freedom. This would therefore appear to be a reflection on the phrasing used in the questionnaire and, if anything seems to show that respondents undertaking the valuation exercise read the descriptions carefully and considered these before valuing the level on the scale.

11.8 Results: Swing-Weighting Exercise

Because of the limited scale of the pilot study, scenarios were not presented in which every capability moved independently. So, with the bottom-up approach scenarios were not presented in which every capability improved independently in turn while all others remained at worst. Likewise with the top-down approach, not every capability was independently at worst while all others remained at best. However, this was undertaken for three of the capabilities: Remaining Physically & Mentally Active, Physical & Mental Well-being, and Social Interaction, and so we will initially look in detail at these three capabilities. Scores are reported in Table 11.7.

Table 11.7: Bottom-up scores with only one Capability at best

Capability	Mean Score	Minimum Score	Maximum Score	Range	n
<i>Bottom-up</i>					
Social Interaction	17.50	5	30	25	6
Physically & Mentally Active	28.33	10	60	50	6
Physical & Mental well-being	42.50	15	70	55	6
<i>Top-down</i>					
Social Interaction	60.00 (40.00)	20 (80)	90 (10)	70	10
Physically & Mentally Active	49.00 (51.00)	30 (70)	80 (20)	50	5
Physical & Mental Well-being	35.00 (65.00)	15 (85)	50 (50)	35	5

The values reported are the 'raw values', the positions at which the scenarios were placed on the scale. So, for example, the scenario in which Social Interaction was at best and the remaining eight capabilities were all at worst

was placed by participants at positions between five and 30 on the scale. The scenario in which Social Interaction was at worst but all of the other capabilities were at best was placed between 20 and 90 on the scale. Values obtained using the bottom-up approach can be interpreted as the gain associated with an improvement in one capability from at worst to at best. In order to calculate the magnitude of shifts down the scale for the top-down approach, i.e. the size of the loss associated with one capability deteriorating from at best to at worst, we must subtract the score from 100; the size of the shift is presented in brackets in Table 11.7.

Table 11.7 shows that while there was a relatively narrow spread of scores for the scenario in which Social Interaction is at best using bottom-up (range = 25), there was a much greater spread of scores for the scenario in which Social Interaction was at worst using top-down (range = 70). Such a large spread in responses could be an indication of a number of phenomena: that respondents interpreted what they were being asked to do in a number of very different ways; that there was little consensus about the relative importance of the capabilities; or that there were wide differences in 'optimism' (the general positive or negative outlook of the respondents).

While respondents would, admittedly, be unlikely to admit that they were baffled by the task, an environment was created in which some respondents felt able to ask for additional guidance on what they were being expected to do and participants appeared satisfied with any additional guidance they were given. Furthermore, in the discussion following the task respondents reported no difficulties and they were able to come up with some justification to support their responses. It would seem then that the large spread of scores was caused in part by disagreement about the relative importance of the capabilities and also the degree to which individual respondents had a positive or negative outlook. So, for example, some respondents reported that losing friendships would be one of the worst things that could happen to them, and so a scenario in which only Social Interaction was at worst was placed a long way down the scale. Others, who were perhaps more independently minded (or even anti-social), reported that Social Interaction

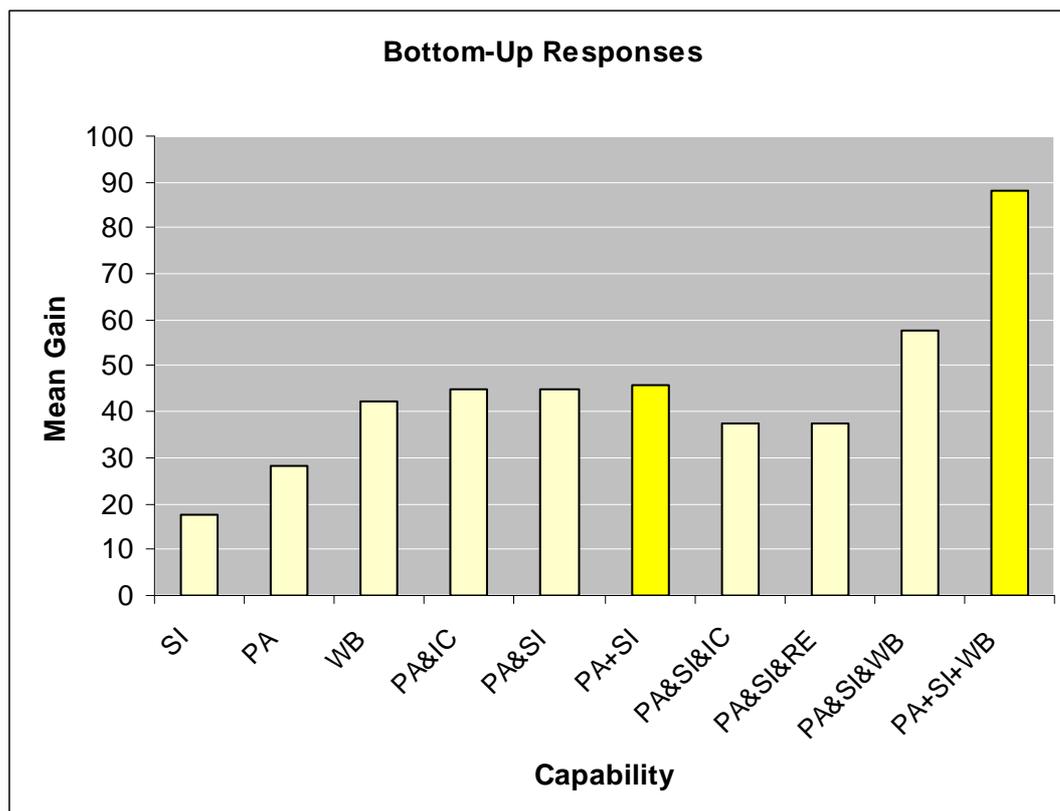
would be less of a loss to them, and that they would be lucky to have all other aspects of life at best.

Those gains which were measured using the bottom-up approach are presented diagrammatically in Figure 11.3. Also displayed in Figure 11.3 are the gains associated with scenarios in which multiple capabilities have improved to at best, from a starting point of all being at worst, i.e. from a starting value of zero. The capabilities which are at best are listed across the x-axis. In each case all other capabilities (i.e. those which are not named) will be at worst.

“PA & SI” represents the scenario in which PA and SI are jointly at best; SI is the capability Social Interaction, WB is Physical and Mental Well-being; RE is Relationships.

“PA + SI” is the sum of the scores from the scenarios in which PA is individually at best and SI is individually at best (these are the darker columns in figure 11.3).

Figure 11.3: Mean Gains Obtained Using Bottom-up



It may be useful, initially, to set out some assumptions about what we would expect to see at the aggregate level. First, we would expect to see scenarios where two or more capabilities are at best dominate scenarios in which only one of those capabilities was at best individually. So, we would expect participants to have scored the scenario in which the capabilities Physically and Mentally Active (PA) and Independence and Control (IC) are jointly at best at least as highly as they did the scenario in which PA was at best individually; some may have scored the scenario in which PA and IC were at best more highly than the scenario in which PA was at best individually. In terms of dominance we would expect to see the following:

1. $PA \& IC \geq PA$
2. $PA \& SI \geq PA$
3. $PA \& SI \geq SI$
4. $PA \& SI \& IC \geq PA \& SI$
5. $PA \& SI \& RE \geq PA \& SI$
6. $PA \& SI \& WB \geq PA \& SI$

Furthermore, if the appropriate functional form for the model is additive, we would expect to see:

7. $PA \& SI = PA + SI$
8. $PA \& SI \& WB = PA + SI + WB$

Finally, we have assumed that the capabilities will not all have equal importance, and so we would expect to see the relationship:

9. $SI \neq PA \neq WB$

In some respects the mean gains expressed in Figure 11.3 appear to be as expected. For example the scenario in which both PA and SI are at best (PA & SI) has been placed at a higher point on the scale than the separate scenarios in which PA and SI are independently at best. Furthermore, if we sum the independent gains associated with PA and SI (PA + SI) we see that the resulting gain is very close to the score given to the scenario in which both PA and SI are at best together (PA & SI); 45.8 compared to 45.

It can also be seen that the capabilities are viewed by respondents as having different degrees of importance, with the score associated with WB being over twice as high as that for SI.

Other aspects of Figure 11.3, however, are not as expected. The mean score for the scenario in which PA and IC were jointly at best and all other capabilities at worst was 45; but, if we consider the scenario in which PA, IC and WB are at best and all others are at worst, then this apparently better scenario has a lower mean score of 37.5. The scenario in which PA, SI and RE are at best is also given a mean score of 37.5, which again is less than the mean score for the scenario in which only PA and SI are best (45).

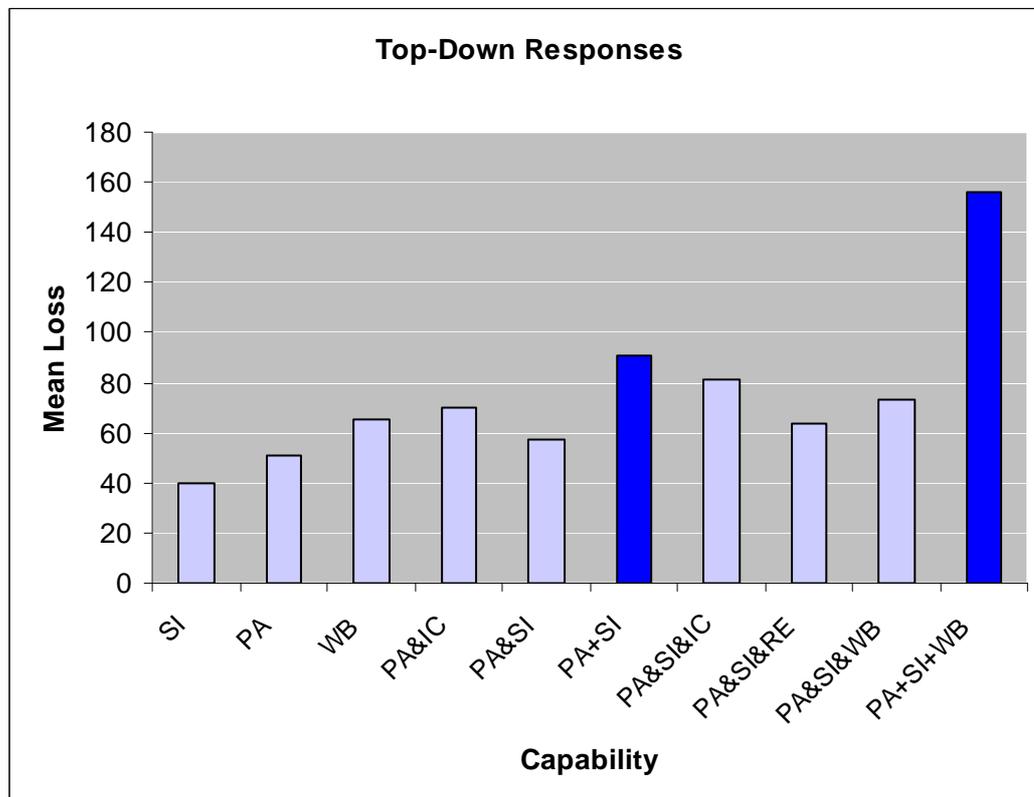
The addition of WB into the scenario does increase the mean score, although this mean score is still significantly less than the sum of the mean scores for the three scenarios in which PA, SI and WB are independently at best, seemingly indicating some significant degree of diminishing gains.

In short, mean responses seem to follow a pattern as would be expected when only one or two capabilities change in any one scenario, but when we have three capabilities changing in a single scenario we see that responses fail to pass a test of dominance and/or that there are significant diminishing gains. While diminishing gains perhaps should have been expected and could be interpreted as being rational, it will be harder to explain the failure to pass a test of dominance. But, before any trend is identified, it will be necessary to consider the mean losses associated with the top-down approach. Mean losses are calculated by subtracting the mean score for a given scenario from 100. Mean losses are presented diagrammatically in Figure 11.4.

Once again, if we look at the capabilities SI, PA and WB, WB appears to be the most important capability of the three, and PA the second most important. This time the mean loss associated with PA and SI being simultaneously at worst (57) is barely higher than the mean loss associated with PA being

independently at worst (51). The sum of the mean losses for the two scenarios in which PA and SI are independently at worst (PA + SI) is 91, which is significantly higher than 57. So, with top-down there appears to be diminishing losses when a scenario contains deterioration in only two capabilities. What is interesting is that although there are these apparent diminishing losses associated with greater numbers of capabilities being at worst, all tests of dominance are met within Figure 11.4.

Figure 11.4: Mean Losses Obtained Using Top-down



Participants considered four scenarios. The first scenario (A) was fixed on the scale at zero (or one hundred), depending on whether they were completing bottom-up or top down. In all cases scenario B involved only one capability changing to at worst (at best), scenario C involved three capabilities changing, and in scenario D two capabilities were changed to be at worst (at best). We would therefore expect to see the following responses for bottom-up (and top-down):

1. $C \geq B$ ($B \geq C$)
2. $C \geq D$ ($D \geq C$)

Scenarios B and D cannot directly be compared as the capabilities being considered were not the same, e.g. Scenario B: SI; Scenario C: SI, IC, PA; Scenario D: IC, PA.

Table 11.8 summarises how many participants gave responses which passed the above tests of dominance, and it can be seen that the overwhelming majority of responses do pass the dominance tests.

Table 11.8: Individual Dominance Tests:

Test	Number of Passes	Percentage Passed
<i>Bottom-Up</i>		
$C \geq B$	12/12	100
$C \geq D$	11/12	91.7
<i>Top-Down</i>		
$B \geq C$	14/15	93.3
$D \geq C$	13/15	86.7

Following discussion in the groups, respondents were offered the chance to give new responses on a new and separate scale. No respondents actually altered the ordering of their responses following discussion and no respondents whose initial responses failed a simple test of dominance chose to change their responses following discussion.

Table 11.9: Relationship between Number of Categories at Best and Score Attached to Scenario

Method	No. At Best	Mean Score	Minimum Score	Maximum Score	Range	n
Top-Down	9	100	-	-	-	-
	8	51.000	15	90	75	20
	7	36.500	10	70	60	10
	6	27.333	5	75	70	15
Bottom-Up	3	47.500	25	80	55	12
	2	41.667	15	75	60	6
	1	29.444	5	70	65	18
	0	0	-	-	-	-

Another aspect of the aggregate data that we can look at is the very simple relationship between the number of capabilities that are at best and the mean score associated with that scenario. This relationship is summarised in Table 11.9.

Table 11.9 shows that there is a very large drop down the scale from the starting point of 100, when all nine capabilities are at best, to 51, when only one capability is at worst. What is apparent is that the shifts up (or down) the scale that are associated with minor changes from the starting point are so large that respondents are leaving only a very small distance on the scale for further shifts associated with further and potentially much more significant changes. It can also be seen that having six of the nine capabilities at best with bottom-down was actually rated lower on the scale (27.3) than having only three at best (47.5) with bottom-up and that likewise, having seven capabilities at best was rated lower than having only two at best. This may, however, have been largely due to the fact that the scenarios in which six and seven capabilities were at best were scored using the top-down method and those in which two and three capabilities were at best were scored using the bottom-up method. Even despite any differences we may expect due to reference point bias, the pattern in Table 11.9 appears to suggest both that respondents over-react to minor changes and, as discussed earlier, that further changes have diminishing effects.

Did allowing people to change their responses following the discussion lead to significantly different results in which the problems outlined above were less apparent? The original dataset was copied exactly, but with new responses entered in the place of old ones. It has been noted already that the majority of respondents chose not to change their original responses, and so actually very few entries in the dataset needed to be changed. And, even after incorporating this small number of altered responses, there was virtually no change to the results.

Let us take a closer look at this phenomenon at a group level. The scenarios presented to the groups, and the mean responses from the groups, are

presented in Table 11.10. Clearly, for the top-down approach the capabilities listed in the scenarios would have been at worst, while all others were at best, and vice-versa for bottom-up. As discussed above, we would expect a scenario in which a particular capability is at worst individually to be preferred to one in which that capability and others are jointly at worst. There is only one group (Group 2) where this was not the case. Responses failing the tests of dominance outlined above are removed and the mean of the remaining responses are presented in Table 11.11.

Table 11.10: Mean Scores by Group:

Group	Scenario B	Score (B)	Scenario C	Score (C)	Scenario D	Score (D)
1 (TD)	SI	54.0	SI, IC, PA	19.0	IC, PA	30.0
2 (BU)	SI	12.5	SI, IC, PA	37.5	IC, PA	45.0
3 (TD)	WB	35.0	WB, SI, PA	27.0	SI, PA	43.0
4 (BU)	WB	38.8	WB, SI, PA	55.0	SI, PA	41.3
5 (TD)	SI	66.0	RE, SI, PA	36.0	PA	49
6 (BU)	SI	50.0	RE, SI, PA	60.0	PA	42.5

Table 11.11: Mean Scores with 'Cleaned Data':

Group	Scenario B	Score (B)	Scenario C	Score (C)	Scenario D	Score (D)
1 (TD)	SI	54.0	SI, IC, PA	19.0	IC, PA	30.0
2 (BU)	SI	5.0	SI, IC, PA	25.0	IC, PA	15.0
3 (TD)	WB	40.0	WB, SI, PA	26.3	SI, PA	47.5
4 (BU)	WB	46.7	WB, SI, PA	66.7	SI, PA	46.7
5 (TD)	SI	62.5	RE, SI, PA	26.3	PA	47.5
6 (BU)	SI	50.0	RE, SI, PA	60.0	PA	42.5

Even in the 'cleaned' data, wherever Well-Being was entered into a scenario it seems to have resulted in the bottom-up scenarios being placed higher on the scale than top-down scenarios, in which less capabilities were 'at worst'. It can also be seen that, as noted at the aggregate level, one individual capability changing to at worst in top-down leads to a fall approximately half way down the scale, a huge drop considering eight other capabilities are still at best. It can also be seen that there is an 'overlap' between the distance moved down the scale when three capabilities change to at worst, and the distance moved up the scale when three capabilities change to at best.

It may be impossible to discover through this study alone why changing just one capability proved to have such a large impact, but four possible explanations are outlined below:

Issues relating to the bottom-up method:

1. The starting position (i.e. all capabilities at worst) is so awful that any improvement will be very significant, and so respondents place scenarios in which there is only a modest improvement at a position that is high on the scale.
2. If one capability is at best (particularly Remaining Active or Well-Being) then respondents find it implausible that the others could simultaneously all be so bad, so they adjust everything else in the scenario to be a little better.

Issues relating to top-down:

3. Well-being is itself simply a hugely important capability, which people expected would affect other aspects of life. One respondent thought that Well-being alone being at worst would be like a 'dark cloud' that would cast a shadow over every other aspect of life, another spoke of well-being deteriorating as triggering a 'domino effect'.

Bottom-up and top-down:

4. Respondents were only asked to consider changes in up to three of the nine capabilities, therefore they perhaps used the full length of the scale without necessarily considering the entire scenario; they are not thinking about the fact that they have not left room on the scale for scenarios in which more than three capabilities change. Respondents are not compensating enough for the fact that 6, 7 or 8 capabilities remain exactly as they were before.

It appears that some of what we are seeing in the results *may* be a form of misspecification bias known as part-whole bias. Part-whole bias is usually seen in contingent valuation studies and is generally thought to occur when "respondents are unable to differentiate between benefit subcomponents or

between the subcomponents and the value for all types of benefits” (Mitchell and Carson 1989, p251). In short, the respondent confuses some broader or wider reaching entity with the entity that the researcher wants to value. What is seen, therefore, is that respondents state much the same willingness to pay for different scales of projects/benefit (Ibid.).

11.8.1 Mean Response Method of Calculating Importance Weights for Capabilities

Peacock *et al.* (2007) use a relatively straightforward method for calculating weights for their attributes, which was outlined in section 10.6. We have three capabilities (PA, SI and WB) for which we can simply apply step one of the process. For example, when only Social Interaction was at best using bottom-up, the mean gain was 17.5; the mean loss when only Social Interaction was at worst in the top-down method was 40, so the importance weight for Social Interaction, using this method, would be 28.75. Peacock *et al.* (2007) state that the purpose of finding the mean of the importance values from bottom-up and top-down methods is to adjust for the potential impact of reference point effects (p906). The weights for these three capabilities are presented in table 10.12. Also contained in Table 11.12 is an importance weight for Independence & Control (IC), which has been calculated using the process outlined in step 2. So, for top-down, the loss associated with PA alone was 51, whereas the loss associated with PA and IC was 70, so for top-down, the importance value for IC is 19.

Table 11.12 Weights for Selected Capabilities by Mean Responses Method:

Capability	Bottom up method	Top down method	Weight
Social Interaction (SI)	17.5	40	28.75
Physically & Mentally Active (PA)	28.3	51	39.65
Physical & Mental Wellbeing (WB)	42.5	65	53.75
Independence & Control (IC)	16.7	19	17.85

Table 11.12 shows that despite IC being reported as being an important capability by respondents in the qualitative work, the importance weight for IC is the lowest of the four and quite a lot smaller than that for SI. This may be an accurate reflection of the importance of IC, but we have no way of checking this, and it is quite possible that the small weight for IC is a function of the method used to arrive at the weight. It was noted earlier that there appeared to be diminishing gains (losses) for scenarios in which increasing numbers of capabilities are at best (worst), and so it seems inevitable that we will get lower importance weights calculated according to step two than we will by using step one and therefore importance weights calculated using these two different methods do not appear to be comparable.

In order to explore this idea further let us first calculate the importance weight for SI as if we didn't have 'independent' importance values for this capability. We see that when PA is independently at best the mean gain is 28.3, and when PA is independently at worst the mean loss is 51. When PA and SI are both at best the mean gain was 45, so the bottom-up importance value, calculated according to step 2, must be 16.7. When PA and SI are both at worst the mean loss was 57, so the top-down importance value for SI must be 6. So, the overall importance weight for SI is 11.35, which is significantly less than 28.75, and would mean that SI would be ranked below PA in terms of importance.

Next, we can explore the outcome of using step two with swings involving more than two capabilities. We can once again calculate the importance weight of IC. With bottom-up, the gain associated with having PA and SI at best is 45, and, as noted earlier, the gain associated with having PA, SI and IC at best is 37.5, giving an importance value for IC of -7.5. The loss associated with having PA and SI at worst is 57, and the loss associated with having PA, SI and IC at worst is 81, giving an importance value of 24. Here then, the importance weight for IC would be 8.25.

It is clear, therefore, that we should have serious concerns about using what has here been termed the Mean Response Method, at least when we have

relatively large numbers of attributes (or in this case capabilities). It would seem sensible to compare importance weights which have all been calculated from 'independent swings' (i.e. through steps one and three) as this way we can be sure that the weights are all comparable. But, if we were only to use gains (losses) associated with scenarios in which just one capability improves (deteriorates) independently then it would seem that we would only be capturing any interaction or any form of trade-off between the capabilities in an extremely limited sense.

11.8.2 Using Regression Analysis to Derive Importance Weights

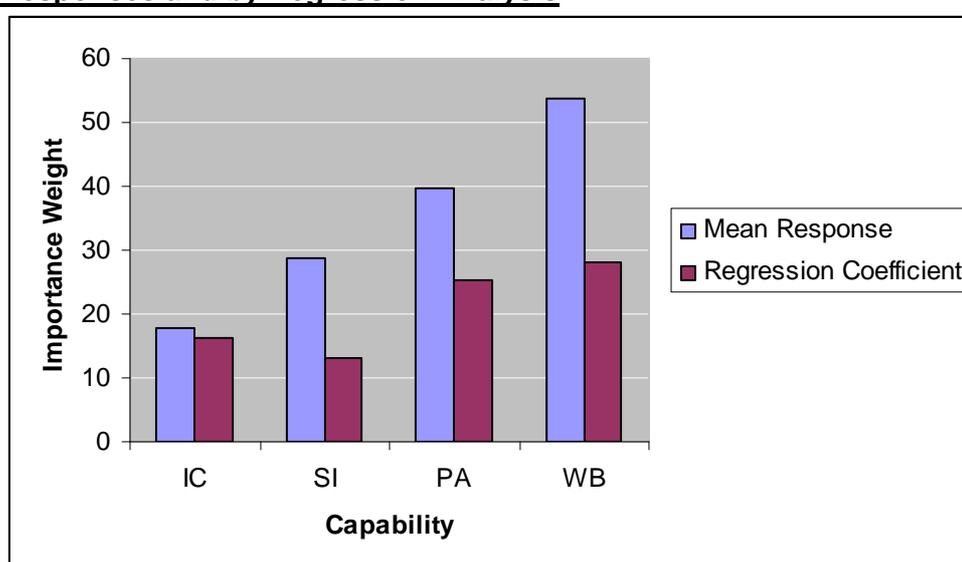
A relatively high proportion of the data collected during the pilot stage will be of no use if the Mean Response Method is used to arrive at the importance weights for capabilities because of the problems, discussed above, of using importance values derived from step two of the Mean Response Method with large numbers of attributes. We have data on the rating scale score for scenarios involving up to three capabilities changing at the same time. One means of deriving importance weights from the whole data set would be to use regression analysis. The dependent variable in the regression model would be the score on the rating scale and the nine capabilities would be included as explanatory variables. Essentially we will have an analysis-of-variance model which will contain nine dummy variables, with each capability being at best, 1, or at worst, 0. Because the intercept will be zero, regression through the origin will be used.

An insufficient number of scenarios were considered by respondents to enable a coefficient to be predicted for every capability, but we can predict the importance weight for the four capabilities discussed previously and these are presented, along with importance weights calculated using the Mean Response method, in Table 11.13. The differences can also be seen in Figure 11.5.

Table 11.13 Comparison of Importance Weights for Capabilities by Mean Responses and by Regression Analysis:

Capability	Weight from Mean Response	Weight from Regression Analysis
Social Interaction (SI)	28.75 (3)	13.12 (4)
Physically & Mentally Active (PA)	39.65 (2)	25.39 (2)
Physical & Mental Wellbeing (WB)	53.75 (1)	28.16 (1)
Independence & Control (IC)	17.85 (4)	16.13 (3)

Figure 11.5: Comparison of Importance Weights for Capabilities by Mean Responses and by Regression Analysis



Depending on which method is used to arrive at the weights, the ordering of the capabilities SI and IC is altered (see Table 11.13). In chapter eight, results were reported from the early stages of the development of the list of higher-order capabilities. As part of this process patients were asked to place the capabilities on a simple visual analogue scale (see Table 9.4). Although the methods used to obtain the VAS scores were very different than the methods used in this phase of the study, it nonetheless provided us with a rank ordering of capabilities, which we can compare to that obtained here (see Table 11.14).

Table 11.14 Comparison of rank ordering of capabilities via three different methods

Rank	Patient Group (Using VAS)	Non-Patient Group (Mean response)	Non-Patient Group (Regression)
1	IC	WB	WB
2	PA	PA	PA
3	WB	SI	IC

It can be seen that there is actually much similarity in the capabilities ranked first, second and third using all three method, although Social Interaction appears out of place, being ranked third in terms of importance using the mean response method. It would be likely that Independence and Control would be ranked third in this column if it were not for the fact that it was not varied independently in the scenarios in Study Phase Three.

For the MAV method used in this phase of the study it would appear that the weights obtained from the raw values according to the procedure used by Peacock et al. (mean responses) are unreliable due to a tendency by respondents to give scenarios in which more than one capability has changed sharply diminishing values. A more reliable alternative appears to be to use regression analysis to compute coefficients for the capabilities, which can be interpreted as weights. But, if the study design could be altered to remove a significant degree of this bias then the mean response method may be useful. An extension to the study was conducted to further explore the presence and possible cause of the bias. This extension is discussed in the following sections.

11.9 Calculating a Capability Score for Respondents

It was suggested in chapter seven that an additive functional form could be suitable for the MAV model in this study. The additive model simply means that the importance weights for the capabilities will sum to unity. The additive function implies that each of the nine capabilities are mutually independent; i.e. the strength of preference over a single capability is unaffected by other, constant, capabilities. This does not appear to be the case, however, given the results presented thus far. Instead, Well-Being appears – as well as being

the capability valued most highly – not to be independent of other capabilities. As has been discussed, the inclusion of Well-Being in a scenario with top-down tended to cause very large shifts down the scale, even if most other capabilities were at best. It was therefore decided to adopt a functional form similar to that used by Peacock and Richardson, where one particular attribute being at its worst level can bring the overall score down to zero (across all attributes). To experience the worst level of capability in terms of physical and mental well-being would be to:

- Never being free of pain,
- Worry about the effects of medication on future health,
- Experiencing severe and frequent side-effects from medication,
- Never being able to sleep well,
- Not being able to get any physical exercise,
- And feel very depressed most of the time.

Intuitively, it is difficult to imagine how when in this state, other aspects of life could continue normally or how pleasure could be taken from other aspects of life.

Weights will be calculated using the same formula as was used by Peacock and Richardson but using values from the regression analysis and not from mean response. As this was a pilot study and we do not have values for all of the nine capabilities, values have been allocated to some capabilities for the purposes of illustration. Where values have been set by the researcher, values were set which appeared to fit with the pattern/range set by the participants and reflect their preferences as stated in the discussion in the groups. This clearly would not be necessary or desirable in a larger study.

From the value (V_i) – which is either the regression coefficient or has been entered by the researcher – the calibrated importance weight is calculated (W_i), where the importance weights sum to 100. The calculation of the importance weight for Self-Respect (W_{SR}) is included below as an example:

$$W_{SR} = [(100 - V_{WB}) / (V_{SR} + V_{SI} + V_{PG} + V_{PA} + V_{ID} + V_{IC} + V_{RE} + V_{EN})] * V_{SR}$$

$$W_{SR} = (71.84 / 120.64) * 13$$

$$W_{SR} = 7.74$$

It should be noted that Well-being is a special case, where $W_{WB} = V_{WB}$.

Table 11.15 contains the calibrated importance weights for all nine of the capabilities. Entries in bold are derived from the responses of participants.

Table 11.15 Importance Weights

Capability	Value (V _i)	Calibrated Weight (W _i)
Self-Respect (SR)	13.00	7.74
Social Interaction (SI)	13.12	7.82
Role of Parent/ Grandparent (PG)	14.00	8.34
Physically & Mentally Active (PA)	25.39	15.12
Identity (ID)	12.00	7.14
Independence & Control (IC)	16.13	9.61
Relationships (RE)	15.00	8.93
Physical & Mental Wellbeing (WB)	28.16	28.16
Enjoyment (EN)	12.00	7.14

The next step is to multiply the values of the levels within functionings by the weight of the corresponding capability. Here, D_{YZ} is the weighted value of level Y on Functioning Z, calculated by multiplying the importance value for level Y on Functioning Z (d_{YZ}) by the calibrated importance weight for Capability i (W_i). The result is divided by 100.

$$D_{YZ} = (d_{YZ} * W_i) / 100$$

The weighted values of the levels within the functionings are listed in Table 11.16. As the best level was always set at 100, this comes out simply equal to the calibrated weight of the corresponding capability. Because the worst level was always set at zero then this still comes out as zero. For now, values for the levels within those functionings corresponding to Well-being are left exactly as they are.

With the weighted responses in the spreadsheet the overall capability score (CS) can then be calculated for each participant. This can be done by taking an average of the weighted scores across the Functionings corresponding to an individual Capability, i . This average will be referred to as DS_i . It is also necessary to take an average of the scores for those Functionings corresponding to Well-being, which will be labelled M . The formula for calculating CS is then:

$$CS = (M / 100) * (W_{WB} + DS_{SR} + DS_{SI} + DS_{PG} + DS_{PA} + DS_{ID} + DS_{IC} + DS_{RE} + DS_{WB} + DS_{EN})$$

Table 11.16: Weighted Values of levels within Functionings (D_{yz}):

Dimension	Best Level	Level A		Level B		Worst Level
		Weighted Mean	Weighted Median	Weighted Mean	Weighted Median	
1A Respected by others	7.740	6.579	6.579	0.227	0.227	0
1B Honest about severity of pain	7.740	6.037	5.418	1.858	1.548	0
1C Believed	7.740	5.132	5.418	2.129	1.935	0
1D Family have accepted illness	7.740	6.192	6.966	2.632	2.322	0
1E Help with everyday tasks	7.740	6.502	6.933	3.096	3.870	0
1F Able to do things worthwhile	7.740	5.805	5.805	1.548	1.548	0
1G Able to earn a wage	7.740	5.031	5.031	-	-	0
2A Meet people through work	7.820	4.497	4.692	-	-	0
2B Go out and socialise	7.820	5.865	5.865	1.564	1.564	0
2C Visit friends & family	7.820	5.521	5.709	1.955	2.346	0
2D Included as part of a social group	7.820	5.474	6.256	1.720	1.564	0
2E Money to go out & socialise	7.820	5.700	5.474	2.057	1.955	0
3A Hug/protect young children	8.340	6.363	6.255	2.711	2.502	0
3B Games/activities with children	8.340	5.421	5.838	2.018	1.668	0
3C Support family financially	8.340	7.089	7.506	1.635	1.668	0
4A Hobbies & Interests	15.120	9.828	9.828	5.670	5.670	0
4B Physically demanding leisure	15.120	11.491	12.096	7.560	7.560	0
4C Paid work	15.120	9.616	10.584	3.644	3.024	0
5A Ask for special treatment	7.140	5.619	5.426	1.485	1.428	0
5B Treated differently	7.140	5.355	5.355	1.614	1.285	0
5C Conversation	7.140	5.569	5.712	2.428	2.856	0
5D Able to disguise/hide pain	7.140	4.820	4.641	1.964	2.142	0

Table 11.16 continued

	Best Level	Level A		Level B		Worst Level
		Weighted Mean	Weighted Median	Weighted Mean	Weighted Median	
6A Self-care	9.610	7.208	7.208	1.922	1.922	0
6B Drive/ use public transport	9.610	6.006	6.247	2.326	1.922	0
6C Walk	9.610	6.727	6.727	3.844	4.805	0
6D Basing routine around medication	9.610	6.804	7.208	1.711	1.826	0
6E Stairs	9.610	5.382	6.727	2.499	2.883	0
6F Ability to plan things in advance	9.610	6.535	6.727	2.883	2.883	0
6G Cook & look after home	9.610	7.111	7.688	3.844	3.844	0
7A(ii) Partner missing out	15.000	12.000	10.800	2.850	3.000	0
7B (ii) Patience & emotional support	15.000	9.870	10.500	2.580	2.250	0
7C Physical Affection	15.000	9.840	9.000	1.350	1.500	0
8A Pain	100	75.6	80.0	13.6	10.0	0
8B Effects of medication on future health	100	70.0	70.0	26.0	30.0	0
8C Side-effects of medication	100	75.0	75.0	30.0	30.0	0
8D Sleep	100	73.3	70.0	29.2	30.0	0
8E Physical exercise	100	76.0	80.0	34.0	40.0	0
8F Depression	100	67.8	68.5	23.3	20.0	0
9A Tiredness & Frustration	12.000	10.200	10.200	2.400	2.400	0
9B Embarrassment	12.000	6.600	6.600	2.400	2.400	0
9C Enjoyment	12.000	10.320	10.800	6.120	6.000	0
9D Concerns about the future	12.000	8.700	8.400	3.300	3.600	0

Using this formula, the best possible score for CS is 100, and the worst is zero. Using the calibrated importance weights, and the importance values for the levels with Functionings, the Capability Score for a respondent ticking level A for every question on the Questionnaire would be 63.814, and the Capability Score for a respondent ticking level B on every question would be 13.316. Notice that the large gap here is due to the way participants in the groups rated the levels within the Functionings, where level B was generally thought to be much worse than level A.

11.10 Part-whole Bias

Part-whole bias was discussed earlier as a problem involving the respondent confusing some broader or wider reaching entity with the entity that the researcher wants to value. What is seen, therefore, is that respondents state much the same willingness to pay (WTP) for different scales of projects/benefit.

It seems unclear from the literature whether the source of this bias should be identified as problems with study design, the inability of respondents to complete the tasks expected of them, or a combination of the two (Boyle, Desvousges et al. 1994).

Baker, Robinson *et al.* (2008) report the results of a literature review of qualitative research conducted alongside contingent valuation studies and report qualitative accounts associated with three studies in which embedding was apparent. Two of these studies related to health. In the first, a study by Currie *et al.*, respondents were asked about their WTP to bring about two expansions in a cancer programme that would treat 300 or 450 additional patients. The authors found insensitivity in the WTP responses to the numbers of cancer patients treated, and in follow-up interviews it was found that respondents were inferring differences between the programmes, such as an assumption that a smaller programme must offer better quality, or that a programme treating fewer patients must be inefficient. Baker *et al.* suggest that respondents “did not perceive that their (collective) contributions would

determine the size of the programme, but were struggling to understand why one programme would be smaller than the other, given a fixed ‘cancer’ budget” (Ibid. 2004, p1437).

In a second study, Shiell and Gold asked respondents what they would be WTP for two vaccines separately and for a combined vaccine delivering both at once. Respondents were then asked to compare the sum of their WTP amounts for the two separate vaccines with that for the combined vaccine in order to explain part-whole bias. Here, respondents felt that the combined vaccine ought to cost less because they were drawing on real life experiences of receiving discounts for buying in bulk, and of items such as 2-in-1 shampoo being cheaper than buying shampoo and conditioner separately. Some respondents also viewed the extra benefit of the combined vaccine as more of a bonus and therefore expected to pay less for the additional protection.

Evidence found and reported by Baker *et al.* suggest that the source of the bias may be the methods by which respondents go about considering their responses, and the ‘baggage’ that they bring into the study, most likely their own real life experiences. Baker *et al.* state that: “the extent to which such anomalies can be ‘designed out’ of WTP studies remains an open question” (p1438). It is acknowledged that although steps can be taken to improve study design, mainly through proper piloting, this may still not fully overcome the problems.

Suggestions can be found in the literature as to how to reduce the extent of part-whole bias. Among general strategies are: inclusion of the larger entity in the scenario, with warning not to confuse the larger entity with that being valued; making the description of the entity more salient, for example through the use of visual aids; have them first value the total resource, even if this is not the subject of the study, and then have them allocate their total WTP amounts for the individual components (Mitchell and Carson 1989).

11.10.1 Amendments to Study Design

A number of possible weaknesses were identified with the original study design. Respondents completing the top-down task were only given a full description of the capabilities at best and then were presented with the scenarios, in which they were told which capabilities had changed and given a new description of only those capabilities. First, it was felt that had respondents been given a description of the worst possible scenario they would have appreciated the huge contrast, and just how serious and undesirable this scenario at the bottom of the scale would be. This may have impressed upon respondents that in the broader scheme of things the scenarios they were considering were relatively 'minor' deviations from the best scenario. Second, there was a concern that respondents were focussing too heavily on the descriptions of the changed capabilities and not fully taking into consideration that many of the other capabilities were still at their best level.

It was felt, and was justified earlier, that scenarios in which more than three capabilities changed simultaneously would become complex, and that limiting the number of capabilities changing in any one scenario would make the task less cognitively demanding. Also, the scenarios were presented in such a way that the first scenario involved only one capability changing, the second involved three that had changed, the third only two. Given the results, it was hypothesised that, were respondents in future groups given an initial scenario in which there was a more dramatic change, then this would help them to 'get a feel' for the full scope of the scale. From such a dramatic initial change, the scenarios which previous respondents had considered would surely appear less significant.

Given a suspicion that what we are seeing in the results so far is a form of part-whole bias, and a reflection on the original study design, the study was extended and a number of changes made to the design and delivery of the Swing-Weighting Exercise, described below.

1. Participants were asked to complete only the Swing-Weighting Exercise (whereas before they started by scaling the levels). But, participants were still talked through an example question and told that the 42 questions are divided into nine categories. It was also still explained that if a respondent completing the questionnaire ticked the top level on every question within a particular category then it would be said that that category would be at best, and vice-versa.
2. Only the top-down method was used.
3. Participants were given a verbal description of the worst possible scenario before they started the task. This was exactly the same as the description of the worst possible scenario given to respondents in the previous bottom-up groups. It was made clear that this scenario was fixed at zero on the scale.
4. Participants were presented with a full written description of every scenario, i.e. a description of every category within the scenario (those at best as well as those at worst).
5. Four scenarios (five including A which was fixed on the scale at 100) were used this time instead of three (four including A). This was simply because with no scaling task more time was available.
6. In the first scenario that respondents had to place on the scale (scenario B) five categories had changed to "at worst". In the initial scenarios a maximum of only three capabilities were allowed to change.

Some changes, such as only asking respondents to complete the swing-weighting task and only using top-down, were made for convenience, to reduce the workload and time needed to run additional groups.

The scenarios presented to respondents were:

- A: All capabilities at best (starting scenario, fixed at 100 on the scale)
- B: EN, WB, RE, PG, PA at worst
- C: RE, PA at worst
- D: RE at worst
- E: PA at worst

11.10.2 Results: Study Extension

The study extension was conducted with three small groups. There were almost equal numbers of male and female respondents and the age of respondents ranged from 20 years up to 65. The sample also included a mix of respondents who were working, studying and retired. Characteristics are summarised in Tables 11.17 and 11.18.

Table 11.17: Characteristics of Participants (Study Three Extension)

Group	ID	Age	Gender	Employment Status
1E	006	48	Male	Employed
1E	A013	20	Male	Student
1E	528	52	Female	Employed
2E	238	36	Female	Other
2E	1017	65	Female	Retired
2E	A014	20	Male	Student
3E	187	44	Female	Employed
3E	937	49	Male	Employed
3E	A015	47	Female	Employed

11.18 Summary of Characteristics of Participants - Study Three Extension:

Characteristic	Frequency	Percentage	Percentage in Initial Groups
Male	4	44.44%	40.74%
Female	5	55.56%	59.26%
Aged 20 - 35	2	22.22%	29.63%
Aged 36 - 50	5	55.56%	18.52%
Aged 51 +	2	22.22%	51.85%
Employed	5	55.56%	29.63%
Retired	1	11.11%	37.04%
Student	2	22.22%	22.22%
Other	1	11.11%	11.11%

Table 11.19 summarises the relationship between the number of capabilities at best and the scores attached to the scenarios. It can be seen that the range of scores given by participants to scenarios in which either eight or seven capabilities were at best was smaller in the study extension than in the main study, 50 and 55 for the extension, respectively, compared to 75 and 60. In Table 11.20 there is a comparison of this relationship in the initial study and in the study extension. Because there were two scenarios (D and E) in which only one capability was at worst (and hence eight were at best) there are 18

observations for this case (an average is presented of all the scores for the two scenarios).

Table 11.19: Relationship between Number of Categories at Best and Score Attached to Scenario for study extension

No. At Best	Mean Score	Minimum Score	Maximum Score	Range	n
9	100	-	-	-	-
8	75.556	50.0	100.0	50.0	18
7	67.778	40.0	95.0	55.0	9
4	34.444	20.0	50.0	30.0	9

Table 11.19 shows that the amendments to the study design resulted in a large reduction in scores for scenarios in which eight and seven capabilities were at best. When only one capability is at worst, then the mean loss was nearer to one quarter (24.4) in the study extension, compared to approximately one half (49) previously. When two capabilities were at worst in the study extension the mean loss was approximately one third (32.2), compared to nearer two-thirds (63.5) previously. Given the small sample size in both groups it would not be meaningful to test whether these differences are statistically significant. With a larger-scale study, it is suggested that an independent samples *t*-test would be suitable.

Table 11.20: Impact of Changes to Study Design

No. At Best	Mean Score (Initial)	Mean Score (Extension)	Difference
9	100	100	-
8	51.000	75.556	-24.556
7	36.500	67.778	-31.278
6	27.333	-	-
4	-	34.444	-
0	0	0	-

Scores for the scenario in which the capability 'Physically and Mentally Active' is at worst in the study extension can be compared to those in which it was at worst in the initial study, this comparison is presented in Table 11.21.

Table 11.21: Top-down scores with only one Capability at worst: Comparison

Capability	Mean Score (Initial)	Range (Initial)	Mean Score (Extension)	Range (Extension)	Difference in Mean Score
Physically & Mentally Active	49.00	50.00	75.56	50.00	-26.56

Again, the difference is striking with a utility loss of 51 in the initial study but only 24 in the study extension. Respondents in the study extension appear to be more sensitive in their valuation to the magnitude of the change in quality of life than did respondents in the initial study, suggesting that alterations to the study design successfully reduced the impact of part-whole bias. It should be realised, however, that the sample size in both phases of the study was small, and that the study was not originally designed to explore this phenomenon. If the MAV model is to be used more extensively in the future then this issue, essentially relating to study design, should first be investigated more fully.

11.11 Summary

This work was only intended to be a pilot study and as such a relatively small number of groups were held (seven) with two to five participants in each. Groups lasted for approximately 90 minutes and consisted of four parts. First, participants were welcomed and told about the background to the study. Then the scaling exercise was explained, an example demonstrated, and the task completed by participants individually. Participants were then asked to discuss the task and their responses as part of a group, with discussion being facilitated by the researchers. Each group was given one of six sets of seven questions for the scaling exercise.

The Swing-weighting exercise was then introduced, an example was given and participants were asked to complete the task individually. Half of the groups were given the top-down version of the swing-weighting exercise and half were given the bottom-up version. Participants started in Scenario A (which was fixed on the scale at either zero or 100) and were then asked to

imagine a further three scenarios (B, C and D). In the scenarios presented to participants the levels on a maximum of three categories (capabilities) were allowed to move.

It has been found that:

1. Results from a swing-weighting exercise suggest that a purely additive functional form for the MAV model would be inappropriate; it appears that we cannot assume independence between the capability 'Well-being' and the other eight capabilities. It was therefore decided to use the same functional form used by Peacock and Richardson where Well-being at worst can reduce the capability score to zero, but there is an additive relationship between the remaining eight capabilities.
2. Part-whole bias appeared to be present in the results from the swing-weighting exercise.
3. Importance weights calculated using steps two and three of the 'mean response' method were found to be markedly smaller than those calculated through steps one and three of that method. When importance weights were instead calculated using regression analysis they appeared to fit better the ranking of capabilities reported by participants in the discussion.

The MAV model has been found to be useful and relatively straightforward to implement, and for respondents to complete. Where problems were encountered they were no more severe than would be expected in most pilot studies, and there is good indication that steps can be taken to resolve them before any wider study might be launched. One such 'issue' was the likely presence of part-whole bias, so far defined and discussed as a phenomenon present within contingent valuation studies (Kahneman and Knetsch 1992). One objection to the use of contingent valuation was methodological issues associated with willingness-to-pay, including the susceptibility to part-whole bias (Svedsater 2000). If it was indeed part-whole bias which was present in the original results then other valuation techniques may be just as susceptible

to this weakness. As suggested earlier, however, initial and tentative findings suggest that part-whole bias can be minimised through careful study design.

An issue that will require consideration whenever multi-attribute health state classification systems are valued – particularly those with a sufficiently large number of possible states so as to prohibit the valuation of each directly – is that of independence between the attributes. It was found in this pilot study that the capability ‘well-being’ cannot be assumed to be mutually independent from the other eight capabilities, and, in reality, still further interaction may be present between the remaining capabilities. While it was possible to adopt a functional form for the MAV model which appears to allow appropriate interaction, the apparent lack of independence caused difficulties during the swing-weighting exercise. Furthermore, had a qualitative element not been included in the study, the interaction would not have been so clearly evident.

Additive functions which have been used in the quality of life literature include the Quality of Well-being scale, the EQ-5D, the SF-6D and the SF-15D Health-related quality of life instrument (Peacock, Richardson et al. 2007). So, for example, in EQ-5D, extreme pain and discomfort and/or extreme anxiety and depression could not independently impact catastrophically on overall quality of life.

There is something of a trade-off between the selection of a functional form that is appropriate given our underlying assumptions, and the use of a simpler and more practical form which may be deemed adequate. Keeney and Raiffa (1993) advise that the additive and multiplicative functions appear to be the more practical for models involving four or more attributes. “Even when the requisite assumptions do not precisely hold over the domains of all the attributes, it may be a good approximation to assume they do, or it may be reasonable to integrate different additive and multiplicative utility functions...” (Ibid. p298). This clearly fits with the handling here of the Capability ‘Well-being’ in the MAV model.

CHAPTER 12: Discussion and Further Recommendations

This thesis has investigated how the outcomes associated with ill health and with health treatments are included within economic evaluations and has assessed the relative merits and weaknesses of various 'standard' approaches, both in general and more specifically in the context of chronic pain. There are three implications of focussing solely on health as the evaluative space in cost-utility analysis, which are: (i) limited scope to compare health interventions with interventions falling under different policy areas; (ii) that focussing only on one aspect of wider quality of life is to omit other useful and important information from the evaluation; (iii) the maximisation of health makes the separation of efficiency and equity impossible. Cost-benefit analysis offers a much broader evaluative space, although there are a number of methodological issues to overcome when designing willingness to pay studies. Economists and philosophers such as Amartya Sen have also questioned how useful utility is as an indication of quality of life or well-being.

There is a vast literature exploring the methodological issues, such as part-whole bias and the effects of budget constraints, associated with willingness to pay and this is one area which will inevitably continue to attract further research. Another area of research attracting increasing attention is the development of alternative approaches to CUA and CBA and experienced utility (or subjective well-being) is an example of one such approach. Capability is a second alternative, in which an entirely different evaluative space is adopted from utility and an attempt is made to define quality of life more objectively and in terms of an individual's ability to achieve various functionings (important beings and doings).

Qualitative work was carried out with chronic pain patients in order to develop a list of nine higher-order capabilities and a total of 42 related capabilities/functionings. Using this list of capabilities as the foundation, a questionnaire was developed and piloted in two stages with two different

groups of patients from the pain clinic at the James Paget University NHS Hospital. The quality of life states defined by the questionnaire were also valued by a sample of the general population using a multi-attribute value method. It has therefore been possible to score responses from the questionnaire and compare these scores to tariff values from the EQ-5D and scores reflecting self-reported health from the EuroQol VAS.

Despite there being a vast literature on the Capability Approach, much of it remains conceptual. There have been suggestions on how to operationalise the approach by authors such as Robeyns (2005b; 2006) and Alkire (2002), and other economists are working to operationalise the approach within the context of health (Anand and Van Hees 2006; Grewal, Lewis et al. 2006; Anand, Hunter et al. 2007; Coast and Horrocks 2007; Coast, Flynn et al. 2008; Lorgelly, Lorimer et al. 2008; Greco, Skordis et al. 2009), but it was inevitable from the outset that much of the work forming this thesis would be exploratory.

This chapter will reflect on: the qualitative methods used; the theoretical issues associated with moving beyond health to quality of life and the wider challenges and implications of adopting a Capability Approach; the usefulness of the MAV method and the issue of part-whole bias; and the policy implications of this research. There will also be recommendations for future research.

12.1 Reflections on the Qualitative Work

Focus groups were conducted with pain patients from the James Paget University NHS Hospital, Norfolk, for which the overall sample was divided into six segments. Segments were defined according to age, gender and employment status. One immediate limitation of the qualitative work was the disappointingly small sample size (n=16). A second limitation was the fact that the patients had all been recruited via one clinic. Although the geographical area which the clinic serves is large, respondents were nonetheless solely from a relatively poor and rural area of one region and were almost entirely of a white British demographic. There were practical

reasons for limiting recruitment in this way, namely that the study was run in partnership with and relied on the support of the clinic staff at the James Paget Hospital, as well as financial constraints. Most importantly, the objective of this work was simply to explore the methods which could be used and to evaluate these methods in order to form recommendations for future research, it was therefore not necessary to recruit across a number of geographical areas.

A laddering questioning technique was used in the focus groups and the qualitative data was analysed using framework analysis. Both of these approaches proved useful and appeared to fit well with the objective of the qualitative work, to arrive at a list of capabilities. Laddering is a questioning technique relying on a series of 'why questions', essentially allowing the researcher to ascertain a respondent's motivation for performing various functionings, or the contribution of these functionings to wider quality of life. Framework analysis involves grouping the type of functionings (sub-themes) discussed in the groups under broader themes, immediately providing us with our higher-order capabilities.

Ideally, the process of undertaking the framework analysis would have been conducted by a research team rather than by an individual. With a team there would have been the scope to discuss and challenge the linkage of sub-themes to broader themes and the interpretation of the meaning of these broader themes, or higher-order capabilities. The fact that the framework analysis was conducted by an individual clearly reflects the nature of a PhD. It is recognised that the 'allocation' or linkage of each of the 42 capabilities/functionings to one of the nine higher-order capabilities will be open to questioning and debate and such scrutiny is necessary. Indeed a degree of scrutiny has been provided through presentation of the final list of higher-order capabilities at forums such as the Health Economists' Study Group and by piloting the questionnaire with a new sample of pain patients from the clinic at the James Paget Hospital.

What was discussed in the focus groups was similar to what has been drawn from qualitative work with pain patients previously and reported in the literature introduced in chapter two. The similarity between the effects of chronic pain on quality of life reported in the literature and the themes identified through qualitative work for this thesis suggests a significant degree of what has been termed by some (Ritchie and Lewis 2003) as 'representative generalisation'. Representative generalisation is the extent to which findings can be inferred to the parent population.

It was noted in chapter eight that there is a degree of similarity between some of the capabilities developed through qualitative work as part of this project and those developed by Grewal *et al.* (2006). Appendix E contains dimensions from the questionnaire developed here for use with pain patients and matches these questions with those developed for use with older people by Grewal, Coast *et al.*, and for use in public health settings by Anand, Lorgelly *et al.* (2008). The similarities between the three sets of capabilities suggests some degree of inferential generalisation (Ritchie and Lewis 2003). Inferential generalisation is a test of whether the findings from a particular study can be generalised, or inferred, to other settings or contexts (Ibid. p264).

Validity is a concept most appropriately associated with quantitative work and it is reported by Ritchie and Lewis that efforts have been made to move away from using the term in a qualitative context, hence the discussion here of representative and inferential generalisation. There is at least, however, an intuitive link here to convergent validity, that constructs which theoretically should be similar are similar; and to external validity, the extent to which the constructs are applicable to other groups within the population and to other contexts or settings (Ibid, p273). The interviews can be thought of as having been a test of face validity, albeit on a very modest scale.

As well as the similarities highlighted above, there were also differences between the three sets of capabilities and this fits with the hypothesis held at

the start of the focus groups that a certain number of capabilities would be context specific.

It was acknowledged in chapter nine that the questions developed through this work are very concentrated on each specific issue identified from the focus groups, and it was suggested that this is a likely reason for the high number of questions that were missed during the piloting stage. The questions on relationships are a good example of this, even to the extent that different questions were asked to those who were single as opposed to those in a relationship. The specific nature of many of the questions, and because of this the need to include such a high volume of questions, is possibly an indication of weakness in the process of analysing the qualitative data, at the stage of identifying and describing the main themes. If the qualitative work was revisited then the number of dimensions (questions) could probably be reduced and perhaps even the number of higher-order capabilities.

The volume of questions as well as their focus and descriptions is likely to have contributed to the volume of missing data at the piloting stage. The list of questions suggested by Anand (which was linked to capabilities originally listed by Nussbaum) has been subjected to a process of piloting and item reduction by Lorgelly *et al.* (2008) which involved the use of factor analysis. One solution would therefore be, like Lorgelly *et al.*, to use factor analysis although it would not be appropriate to proceed with steps such as this before the qualitative data and more specifically the latter steps involved with the framework analysis had been thoroughly reviewed. While there *may* be several questions asking similar things and attracting the same responses it may well be possible to eradicate these by reviewing the process through which they were selected rather than by adding in an extra stage of quantitative work. Factor analysis may then still be a desirable option if on a 'second attempt' the number of questions (dimensions) is still high.

Most questions could also be rewritten with the same meaning but with a slightly broader scope, i.e. to be relevant to a wider population group and encompass a broader range of personal circumstances on the part of the

respondent. Indeed sections such as 'Social Interaction', 'Physically and Mentally Active' and 'Physical and Mental wellbeing', which contain questions more obviously relevant to all groups in society were mostly completed in full. Having presented the lists of capabilities and studied the questionnaire responses, it is felt that the capabilities most in need of consideration are 'Relationships' and 'The Parent and Grandparent Role'. 'Identity' is also similar to the first capability, 'Self-Respect'.

12.2 Theoretical Issues and Contributions

12.2.1 Health Versus Quality of Life

While it would certainly be useful to review the qualitative work, this may not be the sole root of the problem regarding missing data. It was stressed in chapter three that there is no commonly accepted definition of quality of life; indeed what constitutes a good quality of life may well come down in large part to individual preferences, culture and moral values. It may not then be possible to develop a description of quality of life on which all agree and/or identify with. The fact, however, that similar capabilities were identified from work with pain patients to those identified through work with older people and from the writings of Nussbaum suggests that this is not the case.

12.2.2 Lists of Capabilities

It has been noted that there is some overlap between the three lists of capabilities developed in the context of health so far. This suggests that the development of a generic list of capabilities for all patient groups *may* be possible and useful.

It has been Sen's suggestion that the list of capabilities used to form the evaluative space be context specific and both Alkire and Robeyns have suggested adopting a participatory approach to arrive at such a list. It is not clear how broad the interpretation of the context should be. Comparing the list of capabilities developed with pain patients with that developed with older people and with that based on the more philosophical and conceptual writings of Nussbaum suggest that there is at least a core of capabilities important

across all groups. Such Capabilities relate to: love and friendship; independence and control; being able to do things considered to be productive and worthwhile; being free of worry; and being able to fulfil fairly basic daily and recreational activities. It is possible that this list will expand a little as the amount of research conducted in this area expands.

There is scope to develop a generic, or core list in one of two ways: (i) extensive qualitative work with representative samples of the general population, or (ii) the union being found between different sets (lists) of capabilities developed in many different and specific contexts. The latter approach would provide the most accurate picture of the *differences* between the priorities in life of different groups as well as what priorities they share. Those capabilities common to this study and the other two discussed are effectively the union of the sets of capabilities available to us at present within the context of health. What is important is that together with the similarities between the three lists, there are also differences. For example, how older people's considerations of 'love and friendship' have been reported by Grewal *et al.* appears to be distinctly different to the issues surrounding relationships reported by pain patients. It is acknowledged, however, that an element of this perceived difference may be due to analysis and reporting by the two sets of researchers.

Also of importance is the extent to which respondents are reporting their capability and not simply their actual functionings. What is being asked of respondents is whether they *could* do something, not whether they actually do it; whether there is the potential, not whether that potential is realized. Instruments such as the EQ-5D and SF-6D measure functionings, albeit quite a narrow range of health-related functionings. Sen's Capability Approach seeks to achieve equality of opportunity, using capability as the evaluative space. Sen himself has commented that it might not be possible to measure capability, and in fact within the economic development literature the focus tends to be solely on functionings (e.g. the HDI). This has not, however, stopped researchers continuing to give a great deal of time to the debate surrounding the approach, and functionings are still a useful indication of

capability. What is more, the principal motivation for adopting the Capability Approach in this study was to facilitate a move away from measuring health and towards measuring wider quality of life.

A finding that respondents are unable to think in terms of capability and instead respond entirely in terms of their achieved functionings would not then be a damning blow to the capability approach. It would mean though that as researchers we are wasting our time 'dressing' questions up by adding prefixes such as "I am able to...", and "I can have...". From studying the responses to the Capability Questionnaire developed as part of this thesis, it would seem that respondents can and have responded in terms of capability, but that there are boundaries as to how far their 'imagination' will reasonably stretch. When respondents were asked about their ability to do something which it would not be logical for them to do given their circumstances then they often gave no response. Consider as an example from chapter nine, the questions relating to relationships; it was found that respondents who were widowers were likely to skip over this question. This will have implications for the types of questions which can be asked to certain demographic groups. It has already been suggested that the functionings identified within this study should have been made more general for the purposes of the questionnaire.

Grewal, Coast *et al.* (2006) felt that what was important to their respondents was the ability to do things that they considered to be important, and this was their main motivation for adopting the capability approach. Recall from chapter six that Anand and van Hees (2006) believe that their work in developing and using a capability questionnaire demonstrates that it is possible to design questions that distinguish between capabilities and achievements. There is limited evidence, therefore, that people do think and can respond in terms of capability, although this deserves further investigation, possibly through mixed method research (qualitative and quantitative).

12.2.3 Broader Implications of the Research

It was felt that it was appropriate to use a choice-less technique for valuing the capability states defined by the questionnaire. The technique adopted was the multi-attribute value method first explored by Peacock and Richardson (2007) to inform priority setting in the context of mental health services. The MAV method was found to be useful and relatively straightforward both to implement and for respondents to complete. Qualitative work conducted with respondents after they had completed each stage of the MAV method suggested that respondents understood the task, reached their responses based on the information given to them, and had to some extent considered tradeoffs between the capabilities. By tradeoffs it is meant that respondents considered how an improvement in one may help to compensate for the deterioration of another. The process is not complicated by the inclusion of probability as with SG and commonly with WTP, and there were no protest responses as might have been expected with WTP.

What was observed, however, was the suspected presence of part-whole bias, a phenomenon so far only discussed in the WTP literature. Assuming it was part-whole bias causing respondents to 'over-react' in their scoring to relatively small changes in quality of life then there should be a greater awareness of the potential for other types of valuation technique to be susceptible to this form of bias. Findings from the third phase of this study (the valuation of the capability states) suggest that if the susceptibility of valuation techniques other than WTP to part-whole bias is known then steps can be taken at the study design stage to effectively minimise the bias.

A key theme within the feedback from respondents was that the nine capabilities may not have been totally independent. It was commonly felt, for example that if a person experienced the worst level of physical and mental well-being that it was unlikely they could have high achievement on any of the remaining eight capabilities. Once again, revisiting the analysis of the data from the focus groups may lead to a slightly different set of capabilities and would be likely to minimise this problem. At the same time it may never be

possible to have a list of completely independent capabilities, and it may well be the case that the attributes or dimensions on other instruments are also not considered by respondents to be completely independent. Using the EQ-5D as an example, it may be the case that if a person reports either having extreme pain and discomfort or being unable to wash and dress their self then they'd be very unlikely to be able to perform their usual activities with no problems. The MAV method can cope with this lack of independence if there is careful selection of the functional form.

In terms of suggestions for future research, valuing some of the EQ-5D health states using the MAV method may be an interesting exercise. Using a similar functional form as has been adopted for the scoring of the capability states would mean that being in extreme pain on the EQ-5D could reduce the score for the overall health state to zero. This is less dramatic than it first sounds, as a high number of the existing tariff values are below zero, indicating that death would be preferable to some of the health states defined by EQ-5D.

12.3 Policy Implications

One concern relating to the move from health to quality of life, raised at forums such as the Health Economists' Study Group and the International Health Economics Association, is the extent to which wider measures will contain attributes which will actually be affected by changes in health; health being the defining goal of a healthcare system. It may be that broader and more permanent aspects of life dominate to an extent that they mask an improvement in health which would be significant if considered in isolation. Or that it takes longer for an immediate change in health to have a 'knock-on' affect on broader aspects of quality of life. If this is the case then changes in the quality of a patient's life may not be detected during the relatively short time period over which economic evaluations are typically conducted.

Given a full set of scores for the questionnaire, and if a revised version of the questionnaire is used within a larger sample of pain patients, then the sensitivity of results could be compared to the sensitivity found within

instruments such as EQ-5D. This would be an important research question which could be addressed in future research.

The sensitivity of quality of life instruments compared to health instruments is a subtly different argument than the suggestion that health alone should be the defining goal of a healthcare system and that health should therefore be the only output of interest. It was suggested in chapter two that because chronic pain has such a broad impact on a patient's life then an assessment of their health alone would omit outcomes of illness or treatment which are important to the patient. Two further consequences of the insistence that health alone be maximized were the potential to override a patient's opportunity cost (Birch and Donaldson 2003) and the inability to separate efficiency and equity (Coast 2009).

Chronic pain is a complex and prolonged condition. Often there is simply no 'one-off' cure available and so in most cases it will never be observed that the patient is returned to full health. It was noted in chapter two that there are numerous different treatments for pain. Given the range of available treatments, it would seem appropriate for clinicians to base their choice of treatment not only on the severity of the patient's pain but also according to which treatment 'fits' the requirements and lifestyle of the patient, in terms of side-effects, the treatment pathway and a consideration of co-morbidities. In general it seems reasonable to assume that the treatments providing the most effective pain relief will also be those with the most unpleasant side-effects and so some balance or trade-off is required between achieving pain relief and the patient being able to function in terms of basic everyday life. It would seem likely that in reaching a decision the clinician would consider non-health outcomes. Having said this, given the definition of health used by the World Health Organisation, it may be difficult to identify an aspect of quality of life which is unrelated to health.

Linked to the fact that chronic pain has a prolonged and often significant affect on the patient's everyday life, is the fact that they are likely to need and receive support from the state which is provided by agencies other than the

healthcare system. In the introduction to this thesis it was noted that there is a desire by decision makers to deliver 'joined up' services and to bring consistency to the process for evaluating the cost-effectiveness of different interventions. The government's desire to deliver joined up services appears to indicate recognition that there can be overlap in the benefits delivered by different agencies, and that the overall benefit when agencies work together can be greater than the sum of what is achieved by the agencies working alone. Intuitively, the sum of these efforts must represent something broader than improvement in health alone. As long as health is used as the evaluative space in economic evaluations of treatments provided by the health care system then comparisons of the cost-effectiveness of health interventions with interventions provided by 'partner' agencies will always be difficult or even impossible.

Although the questionnaire developed as part of this study should be able to encompass the impact of support from outside of the health sector such as from social services, it is also condition specific. If every list of capabilities is different because it has been developed with a different and specific population group then the capability approach would be as restrictive as cost-utility analysis. There is only scope to use the Capability Questionnaire developed here to compare interventions (albeit by any agency) which are targeted at chronic pain patients. Recall that Sen has spoken of the need for lists of capabilities to be context specific. It is not clear how broadly the term 'context' should be defined. It was suggested earlier in this chapter, however, that the overlap between the different lists of capabilities identified so far suggests it may be possible to arrive at one core set of generic capabilities. This list would be developed for use in the context of evaluating the cost-effectiveness of public sector activities in the UK.

12.4 Conclusion

The capability approach has been used as a framework in which to develop an outcome measure which encompasses quality of life rather than health or health-related quality of life. Although it has been suggested that the

qualitative work undertaken should be reviewed, the key themes identified broadly match those which have been identified through previous qualitative studies involving patients with chronic pain.

It is unclear to what extent respondents actually think in terms capability (in terms of the opportunities they have), and this will be a useful and interesting focus for future research. Concerns had been raised that an instrument with a broader focus (assessing overall quality of life) may be insensitive to changes in health alone; evidence from this study implies that this is not the case.

The capability states defined by the questionnaire have been valued using a multi-attribute value method; a choice-less technique which was found to be simple both to operationalise and for respondents to complete. Although issues such as budget constraints, the complication of including probabilities and protest votes by respondents unwilling to value health in monetary terms are all avoided, it was suspected that part-whole bias was affecting the results. As part-whole bias is a phenomenon which, so far, has been largely confined to the literature on willingness to pay, the presence of the bias here should be taken as a warning that other valuation techniques may not be immune. Steps were taken to minimise the bias and these proved to be successful, suggesting if the susceptibility of valuation techniques other than WTP to part-whole bias is known then steps can be taken at the study design stage to effectively minimise the bias.

The nine higher-order capabilities are similar to capabilities developed through qualitative work conducted with older people, and from a more conceptual list drawn up by Martha Nussbaum. Evidence at this stage therefore suggests that it will be possible to develop one core set of capabilities, against which will be relevant and appropriate to assess the impact of public sector interventions and services for inclusion in economic evaluations. The significance of this is that health services will be able to be evaluated alongside social services and other public sector projects, something which is not possible under the current preferred approach, cost-utility analysis.

References

- Aberg, A. C., B. Sidenvall, et al. (2005). "On loss of activity and independence, adaptation improves life satisfaction in old age - a qualitative study of patients' perceptions." QUALITY OF LIFE RESEARCH **14**(4): 1111-1125.
- Alkire, S. (2002). Valuing Freedoms. Sen's Capability Approach and Poverty Reduction. New York, Oxford University Press.
- Alkire, S. (2005). "Why The Capability Approach?" Journal of Human Development **6**(1): 115-133.
- Alkire, S. (2005b). "Subjective Quantitative Studies of Human Agency." Social Indicators Research **74**: 217-260.
- Anand, P. (2005). "QALYs and Capabilities: a comment on cookson." Health Economics **14**: 1283-1286.
- Anand, P., G. Hunter, et al. (2007). The Measurement of Capabilities.
- Anand, P., G. Hunter, et al. (2005). "Capabilities and Well-Being: Evidence Based on the Sen-Nussbaum Approach to Welfare." Social Indicators Research **74**(1): 9-55.
- Anand, P. and M. Van Hees (2006). "Capabilities and Achievements: An Empirical Study." The Journal of Socioeconomics **35**: 268-284.
- Arnold, L. M., D. L. Goldenberg, et al. (2007). "Gabapentin in the Treatment of Fibromyalgia: A Randomized, Double-Blind, Placebo-Controlled, Multicenter Trial." Arthritis & Rheumatism **56**(4): 1336-1344.
- Arrow, K. (1963). "UNCERTAINTY AND THE WELFARE OF ECONOMICS OF MEDICAL CARE." The American Economic Review **53**(5): 941-973.
- Baker, R. and A. Robinson (2004). "Responses to standard gambles: are preferences 'well constructed'?" HEALTH ECONOMICS **13**(1): 37-48.
- Baker, R., A. Robinson, et al. (2008). "How do respondents explain WTP responses? A review of the qualitative evidence."
- Bala, M. V., L. L. Wood, et al. (1998). "Valuing Outcomes in Health Care: A Comparison of Willingness to Pay and Quality Adjusted Life Years." Journal of Clin Epidemiol **51**(8): 667-676.
- Baldwin, S., C. Godfrey, et al. (1990). Quality of Life: Perspectives & Policies. London, Routledge.
- Bergman, S. (2005). "Psychosocial aspects of chronic widespread pain and fibromyalgia." DISABILITY AND REHABILITATION **27**(14): 675-683.
- Biggeri, M., R. Libanora, et al. (2006). "Children Conceptualising their Capabilities: Results of a Survey Conducted During the First Childrens' World Congress on Child Labour." Journal of Human Development **1**: 59-83.
- Birch, S. and C. Donaldson (2003). "Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism?" Social Science and Medicine.

- Blomqvist, K. and A. Edberg (2002). "Living with persistent pain: experiences of older people receiving home care." Journal of Advanced Nursing **40**(3): 297-306.
- Boorse, C. (1977). "Health as a Theoretical Concept." Philosophy of Science.
- Borell, L., E. Asaba, et al. (2006). "Exploring experiences of "participation" among individuals living with chronic pain." Scandinavian Journal of Occupational Therapy **13**: 76-85.
- Bowling, A. (1997). Measuring Health: A review of Quality of Life Measurement Scales. Maidenhead, Open University Press.
- Boyle, K., W. Desvousges, et al. (1994). "An Investigation of Part-Whole Biases in Contingent Valuation Studies." Journal of Environmental Economics and Management **27**: 64-83.
- Brazier, J., J. Ratcliffe, et al. (2007). Measuring and Valuing Health Benefits for Economic Evaluation. Oxford, Oxford University Press.
- Brazier, J. (2007). "'Is there a case for using visual analogue scale valuations in CUA' by Parkin and Devlin - A response: 'Yes there is a case, but what does it add to ordinal data?'" HEALTH ECONOMICS **16**(6): 645-647.
- Brazier, J., J. Roberts, et al. (2002). "The estimation of a preference-based measure of health from the SF-36." Journal of Health Economics **21**(2): 271.
- Brazier, J., J. Roberts, et al. (2004). "A comparison of the EQ-5D and SF-6D across seven patient groups." HEALTH ECONOMICS **13**(9): 873-884.
- Breen, J. (2002). "Transitions in the Concept of Chronic Pain." Advanced Nursing Science **24**(4): 48-59.
- Brooks, R. (1996). "EuroQol: the current state of play." Health Policy **37**(1): 53-72.
- Broome, J. (1978). "Trying to Value a Life." Journal of Public Economics **9**: 91-100.
- Broome, J. (1985). "The Economic Value of Life." Economica **52**(27): 281-294.
- Broome, J. (2007). Can there be a preference-based utilitarianism? Justice, Political Liberalism and Utilitarianism: Themes from Harsanyi and Rawls. M. Fleurbaey, M. Salles and J. Weymark, Cambridge University Press.
- Brouwer, W., A. J. Culyer, et al. (2008). "Welfarism vs. extra-welfarism." Journal of Health Economics **27**: 325-338.
- Brouwer, W. B. F. and M. A. Koopmanschap (2000). "On the Economic Foundations of CEA: Ladies and Gentlemen, Take Your Positions!" Journal of Health Economics **19**(4): 439.
- Carr, A. J., B. Gibson, et al. (2001). "Measuring Quality of Life: Is Quality of Life Determined by Expectations or Experience?" BMJ **322**: 1240-1243.
- Carson, M. G. and G. J. Mitchell (1998). "The Experience of Living with Persistent Pain." Journal of Advanced Nursing **28**(6): 1242-1248.
- Carter, B., K. Lambrenos, et al. (2002). "A pain workshop: an approach to eliciting the views of young people with chronic pain." Journal of Clinical Nursing **11**: 753-762.

- Chowdhury, S. and L. Squire (2006). "Setting weights for aggregate indices: An application to the commitment to development index and human development index." Journal of Development Studies **42**(5): 761-771.
- Chuck, A., W. Adamowicz, et al. (2009). "The Willingness to Pay for Reducing Pain and Pain-Related Disability." VALUE IN HEALTH **12**(4): 498-506.
- Clark, D. A. (2005). "Sen's Capability Approach and the Many Spaces of Human Development." Journal of Human Development **8**: 1339-1368.
- Coast, J. (2004). "Is economic evaluation in touch with society's health values?" British Medical Journal **329**(7476): 1233.
- Coast, J. (2009). "Maximisation in extra-welfarism: A critique of the current position in health economics." Social Science & Medicine **69**: 786-792.
- Coast, J., T. Flynn, et al. (2006). Developing an Index of Capability for Older People: A New Form of Measure for Public Health Interventions? Future of Health Conference. Cambridge.
- Coast, J., T. Flynn, et al. (2005). Developing an Index of Capability for Older People: A new Form of Measure for Public Health Interventions? Health Economics Study Group. Cambridge.
- Coast, J., T. Flynn, et al. (2008). "Valuing the ICECAP capability index for older people." Social Science & Medicine **67**: 874-882.
- Coast, J. and S. Horrocks (2007). "Developing attributes and levels for discrete choice experiments using qualitative methods." Journal of Health Services Research Policy **12**(1): 25 - 30.
- Coast, J., T. Peters, et al. (2008). "An Assessment of the Construct Validity of the Descriptive System for the ICECAP Capability Measure for Older People." Qualitative Life Research **17**: 967-976.
- Coast, J., R. D. Smith, et al. (2007a). It's Not Winning That Counts Its The Taking Part: Should We Be Assessing Capability Rather Than Achievement? Health Economists' Study Group. Birmingham.
- Coast, J., R. D. Smith, et al. (2008). "Welfarism, extra-welfarism and capability: The spread of ideas in health economics." Social Science & Medicine **67**: 1190-1198.
- Coast, J., R. D. Smith, et al. (2008a). "Should the capability approach be applied in health economics?" HEALTH ECONOMICS **17**: 667-670.
- Coast, J., R. D. Smith, et al. (2008c). "The influence of capabilities on health care decision making in the UK." Social Science & Medicine **67**: 1190-1198.
- Coniam, S. W. and A. W. Diamond (1994). Practical Pain Management. Oxford, Oxford Medical Publications.
- Cookson, R. (2005). "QALYs, and the capability approach." HEALTH ECONOMICS **14**(8): 817-829.
- Cookson, R. (2005b). "QALY's and Capabilities: a response to Anand." HEALTH ECONOMICS **14**: 1287-1289.
- Cooper, N. (2000). The Economic Impact of Rheumatoid Arthritis In The First Five Years. School of Medicine, Health Policy & Practice. Norwich, University of East Anglia.
- CPPC. "Chronic Pain Policy Coalition: "Chronic Pain in the UK"." Retrieved 10 March, 2008, from <http://www.paincoalition.org.uk/challenge.html>.

- Culyer, A. J. (1991). *The Normative Economics of Health Care Finance and Provision. Providing Healthcare: The Economics of Alternative Systems of Finance and Delivery*. A. McGuire, P. Fenn and K. Mayhew. Oxford, Oxford University Press.
- Culyer, A. J. and R. G. Evans (1996). "Mark Pauly on welfare economics: Normative rabbits from positive hats." *Journal of Health Economics* **15**: 243-251.
- Damschroder, L. J., B. J. Zikmund-Fisher, et al. (2005). "The impact of considering adaptation in health state valuation." *SOCIAL SCIENCE & MEDICINE* **61**(2): 267-277.
- Dewar, A., M. White, et al. (2003). "Using nominal group technique to assess chronic pain, patients' perceived challenges and needs in a community health region." *Health Expectations* **6**: 44-52.
- Dolan, P. (2001). *Output Measures and Valuation in Health. Economic Evaluation in Health Care: Merging theory with practice*. M. Drummond and A. McGuire. Oxford, Oxford University Press.
- Dolan, P. (2008c). "Developing methods that really do value the 'Q' in the QALY." *HEALTH ECONOMICS* **3**: 69-77.
- Dolan, P. and D. Kahneman (2008). "Interpretations of utility and their implications for the valuation of health." *The Economic Journal* **118**: 215-234.
- Dolan, P. and T. Peasgood (2008). "Measuring Well-being for Public Policy: Preferences or Experiences?" *Journal of Legal Studies* **37**(2): S5-S31.
- Dolan, P. and A. Robinson (2001). "The Measurement of Preferences Over the Distribution of Benefits: The Importance of the Reference Point." *European Economic Review* **45**: 1697-1709.
- Dolan, P. and A. Tsuchiya (2009). "The Social Welfare Function and Individual Responsibility: Some Theoretical Issues and Empirical Evidence." *Journal of Health Economics* **28**: 210-220.
- Donaldson, C. (1999). "Valuing the benefits of publicly-provided health care: does 'ability to pay' preclude the use of 'willingness to pay'?" *Social Science & Medicine*.
- Drummond, M. and A. McGuire (2001). *Economic Evaluation in Health Care*. Oxford, Oxford University Press.
- Drummond, M., M. Sculpher, et al. (2005). *Methods for the Economic Evaluation of Health Care Programmes*. New York, Oxford University Press.
- Dysvik, E., T. C. Lindstrom, et al. (2004). "Health-Related Quality of Life and Pain Beliefs Among People Suffering From Chronic Pain." *Pain Management Nursing* **5**(2): 66-74.
- Finch, H. and J. Lewis (2003). *Focus Groups. Qualitative Research Practice: A Guide for social Science Students and Researchers*. J. Ritchie and J. Lewis. London, Sage.
- Fink, A. (1995). *The Survey Handbook*. London, SAGE Publications.
- Gaertner, W. and X. Yongsheng (2006). "Capability Sets as the Basis of a New Measure of Human Development." *Journal of Human Development* **7**(3): 311-321.

- Gamlin, R. and T. Lovel (2002). Pain Explained: A guide for patients and carers. St Albans, Altman.
- Gibbs, A. (1997). Social Research Update, University of Surrey. **Winter, Issue 19**.
- Glendinning, C., B. Dowling, et al. (2005). "Partnerships between health and social care under 'New Labour': smoke without fire? A review of policy and evidence." Evidence & Policy **1**(3): 365-381.
- Gold, M., J. Siegel, et al. (1996). Cost-Effectiveness in Health and Medicine. New York, Oxford University Press.
- Goossens, M., M. Van Molken, et al. (1998). "Health economic assessment of behavioural rehabilitation in chronic low back pain: A randomised clinical trial." HEALTH ECONOMICS **7**(1): 39-51.
- Goossens, M., J. W. S. Vlaeyen, et al. (1999). "Patient utilities in chronic musculoskeletal pain: how useful is the standard gamble method?" PAIN **80**(1-2): 365-375.
- Goossens, M. E. J. B., M. Rutten-van Molken, et al. (1996). "Cognitive-educational treatment of fibromyalgia: A randomized clinical trial. II. economic evaluation." The Journal of Rheumatology **23**(6): 1246-1254.
- Greco, G., J. Skordis, et al. (2009). Methodological challenges for developing an outcome measure based on Sen's Capability Approach for assessing women's wellbeing in rural Malawi. Health Economists' Study Group. Sheffield.
- Greiner, C. Bug, et al. (2006). "Economic evaluation of Durogesic in moderate to severe, nonmalignant, chronic pain in Germany." Eur J Health Econ **7**(4): 290-296.
- Grewal, I., J. Lewis, et al. (2006). "Developing attributes for a generic quality of life measure for older people: Preferences or Capabilities?" Social Science & Medicine **62**: 1891-1901.
- Groot, W. (2000). "Adaptation and scale of reference bias in self-assessments of quality of life." Journal of Health Economics **19**(3): 403.
- Haas, M., R. Sharma, et al. (2005). "Cost-Effectiveness of Medical and Chiropractic Care for Acute and Chronic Low Back Pain." Journal of Manipulative and Physiological Therapeutics.
- Hadjimichael, O., R. D. Kerns, et al. (2007). "Persistent Pain and Uncomfortable Sensations in Persons with Multiple Sclerosis." Pain **127**: 35-41.
- Haefeli, M., A. Elfering, et al. (2008). "A Cost-Benefit Analysis Using Contingent Valuation Techniques: A feasibility Study in Spinal Surgery." VALUE IN HEALTH **11**(4): 575-588.
- Hammitt, J. K. and J. D. Graham (1999). "Willingness to pay for health protection: Inadequate sensitivity to probability?" Journal of Risk and Uncertainty **8**: 33-62.
- Harris, S., S. Morley, et al. (2003). "Role loss and emotional adjustment in chronic pain." PAIN(1-2): 2):.
- HM Government (2006). Our health, our care, our say: a new direction for community services. Department of Health, http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4127459.pdf.

- HM Government (2009). Shaping the Future of Care Together, http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_102732.pdf.
- HM Treasury The Green Book: Appraisal and Evaluation in Central Government, http://www.hm-treasury.gov.uk/d/green_book_complete.pdf.
- House of Commons Health Committee (2008). National Institute of Health and Clinical Excellence: First Report of Session 2007-2008, The Stationary Office Limited. **1**.
- Hurley, J. (2001). "Commentary on "Health-related quality-of-life research and the capability approach of Amartya Sen"." QUALITY OF LIFE RESEARCH **10**(1): 57-58.
- Jefferson, T., V. Demicheli, et al. (2000). Elementary Economic Evaluation In Health Care. London, BMJ Books.
- Kahneman, D. and J. L. Knetsch (1992). "Valuing Public Goods: The Purchase of Moral Satisfaction." Journal of Environmental Economics and Management **22**: 57-70.
- Kaplan, G. and O. Baron-Epel (2003). "What lies behind the subjective evaluation of health status?" Social Science & Medicine **56**(8): 1669-1676.
- Kaplan, R. M., J. P. Anderson, et al. (1998). "The Quality of Well-Being Scale: critical similarities and differences with SF-36." Int J Qual Health Care **10**(6): 509-520.
- Kahneman, D. and A. Krueger (2006). "Developments in the Measurement of Subjective Well-being." Journal of Economic Perspectives **20**(1): 3-24.
- Kahneman, D., B. U Ca, et al. (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem." Journal of Political Economy **98**(6): 1325.
- Keeney, R. L. and H. Raiffa (1993). Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge, Cambridge University Press.
- Kemler, M. A. and C. A. Furnee (2002). "Economic evaluation of spinal cord stimulation for chronic reflex sympathetic dystrophy." NEUROLOGY **59**(8): 1203-1209.
- Kerr, S., G. Fairbrother, et al. (2004). "Patient characteristics and quality of life among a sample of Austalian chronic pain clinic attendees." Internal Medicine Review **34**: 403-409.
- Khan, F. and J. Pallant (2007). "Chronic pain in multiple sclerosis: Prevalence, characteristics, and impact on quality of life in an Australian community Cohort." JOURNAL OF PAIN **8**(8): 614-623.
- Khanna, D., M. Ahmed, et al. (2008). "Willingness to Pay for a Cure in Patients with Chronic Gout." MEDICAL DECISION MAKING **28**: 606-613.
- Kind, P. (1990). Issues in the Design and Construction of a Quality of Life Measure. Quality of Life: Perspectives and Policies. S. Baldwin, C. Godfrey and C. Propper. London, Routledge.

- Kinghorn, P., A. Robinson, et al. (2009). Can we value 'capability'?: Findings from a pilot study applying a Multi-Attribute Value Method. Health Economists' Study Group. Manchester.
- Kinghorn, P. and R. Smith (2008). From Theory to Practice: Are we capable of operationalising the Capability Approach. Health Economists' Study Group. Norwich.
- Kitzinger, J. (1994). "The Methodology of Focus Groups: The importance of interation between research participants." Sociology of Health & Illness **16**: 103-121.
- Klasen, S. (1997). "Poverty, Inequality and Deprivation in South Africa: An Analysis of the 1993 SALDRU Survey." Social Indicators Research **41**: 51-94.
- Klasen, S. (2000). "Measuring Poverty and Deprivation in South Africa." REVIEW OF INCOME AND WEALTH **46**(1): 33-58.
- Kothe, R., T. Kohlmann, et al. (2007). "Impact of low back pain on functional limitations, depressed mood and quality of life in patients with rheumatoid." Pain.
- Kuklys, W. (2005). Amartya Sen's Capability Approach: Theoretical Insights and Empirical Applications. Berlin, Springer.
- Kumar, K., G. Hunter, et al. (2002). "Treatment of chronic pain by using intrathecal drug therapy compared with conventional pain therapies: a cost-effectiveness analysis." Journal of Neurosurgery **97**(4): 803-810.
- Kwan, K. S. H., L. J. Roberts, et al. (2005). "Sexual dysfunction and chronic pain: the role of psychological variables and impact on quality of life." European Journal of Pain.
- Lame, I. E., L. P. Madelon, et al. (2005). "Quality of life in chronic pain is more associated with beliefs about pain, than with pain intensity." EUROPEAN JOURNAL OF PAIN **9**: 15-24.
- Lancsar, E. and J. J. Louviere (2008). "Conducting Discrete Choice Experiments to Inform Healthcare Decision Making: A User's Guide." PHARMACOECONOMICS **26**(8): 661-677.
- Layard, R. (2006). "Happiness and public policy: a challenge to the profession." The Economic Journal **116**: C24-C33.
- Loeser, J. D. and R. Melzack (1999). "Pain: an overview." Lancet **353**(9164): 1607-1609.
- Loomes, G. (2006). "(How) Can we value health, Safety and the Environment?" Journal of Economic Psychology **27**: 713-736.
- Loomes, G. and L. McKenzie (1989). "The use of QALYs in health care decision making." Social Science and Medicine.
- Lords Economic Affairs Committee (2006). Economic Affairs - Fifth Report, www.parliament.uk.
- Lorgelly, P., K. Lorimer, et al. (2008). The Capability Approach: developing an instrument for evaluating public health interventions, University of Glasgow.
- Lorimer, K., P. Lorgelly, et al. (2007). Using mixed methods to operationalise the Capability Approach: an application in public health. HDCA Conference.

- Louviere, J. J., D. Hensher, et al. (2000). Stated Choice Methods: Analysis and Application. Cambridge, Cambridge University Press.
- Ludbrook, A. and D. Cohen (2003). Economics of Health and Health Improvement. Advances in Health Economics. A. Scott, A. Maynard and R. Elliott. Chichester, Wiley: 223-240.
- Mazumdar, K. (1999). "Measuring the well-beings of the developing countries: Achievement and improvement indices." Social Indicators Research **47**: 1-60.
- Megone, C. (1990). The Quality of Life: Starting from Aristotle. Quality of Life: Perspectives and Policies. S. Baldwin, C. Godfrey and C. Propper. London, Routledge.
- Meyer-Rosberg, K., C. Burckhardt, et al. (2001). "A Comparison of the SF-36 and Nottingham Health Profile in Patients with Chronic Neuropathic Pain." EUROPEAN JOURNAL OF PAIN **5**: 391-403.
- Miles, S. and G. Rowe (2004). The Laddering Technique. In: "Doing Social Psychology Research". G. M. Breakwell (Ed). Oxford, BPS Blackwell.
- Mitchell, R. C. and R. Carson (1989). Using Surveys to Value Public Goods. Washington, D. C., Resources for the Future.
- Mooney, G. and E. Russell (2003). Equity in Health Care: The Need for a New Economics Paradigm? Advances in Health Economics. A. Scott, A. Maynard and R. Elliott. Chichester, Wiley: 205-221.
- Mooney, G. and J. A. Olsen (1991). QALYs: Where next? Providing Health Care: The Economics of Alternative Systems of Finance and Delivery. A. McGuire, P. Fenn and K. Mayhew. New York, Oxford University Press.
- Moons, P. (2004). "Why call it health-related quality of life when you mean perceived health status?" European Journal of Cardiovascular Nursing **3**: 275-277.
- Morgan, D. L. (1996). "Focus Groups." Annual Review of Sociology **22**: 129-152.
- Morgan, D. L. (1997). Focus Groups As Qualitative Research. California, SAGE.
- Morris, M. D. (1978). "A physical quality of life index." Urban Ecology.
- Morris, S., N. Devlin, et al. (2007). Economic Analysis in Health Care. Chichester, Wiley.
- MVH Group (1995). The Measurement and Valuation of Health: Final Report on the modelling of Valuation Tariffs, Centre for Health Economics, University of York.
- MVH.Group (1995). The Measurement and Valuation of Health: Final Report on the modelling of Valuation Tariffs, Centre for Health Economics, University of York.
- Neighbors, D. M., T. J. Bell, et al. (2001). "Economic evaluation of the fentanyl transdermal system for the treatment of chronic moderate to severe pain." JOURNAL OF PAIN AND SYMPTOM MANAGEMENT **21**(2): 129-143.
- NICE (2008). Guide to the methods of technology appraisal, National Institute for Health and Clinical Excellence.
- NICE. (2009). "About NICE." Retrieved December 2009, from <http://www.nice.org.uk/aboutnice/>.

- Niemisto, L., P. Rissanen, et al. (2005). "Cost-effectiveness of combined manipulation, stabilizing exercises, and physical consultation compared to physical consultation alone for chronic low back pain: A prospective randomized trial with 2-year follow-up." *Spine* **30**(10): 1109-1115.
- Nussbaum, M. and A. Sen (1993). *The Quality of Life*. Oxford, Clarendon Press.
- Nussbaum, M. C. (2003). "Capabilities as Fundamental Entitlements: Sen and Social Justice." *Feminist Economics* **9**(2-3): 33.
- Oliver, A. J. (2004). "Prioritising health care: is "health" always an appropriate maximand?" *MEDICAL DECISION MAKING* **24**(3): 272-280.
- Olsaretti, S. (2005). "Endorsement and freedom in Amartya Sen's capability approach." *ECONOMICS AND PHILOSOPHY* **21**(1): 89-108.
- Olsen, J. A. and R. D. Smith (2001). "Theory versus practice: a review of 'willingness-to-pay' in health and health care." *Health Econ* **10**(1): 39-52.
- ONS. (14 November 2008). "Office of National Statistics: Labour Market Data." Retrieved June 2009, from <http://www.statistics.gov.uk/cci/nugget.asp?id=285>.
- Panigrahi, R. and S. Sivramkrishna (2002). "An Adjusted Human Development Index: Robust Country Rankings with Respect to the Choice of Fixed Maximum and Minimum Indicator Values." *Journal of Human Development* **3**: 301-311.
- Paulson, M., E. Danielson, et al. (2002). "Struggling for a Tolerable Existence: The Meaning of Men's Lived Experiences of Living with Pain of Fibromyalgia Type." *Qualitative Health Research* **12**(2): 238-249.
- Peacock, S. J., J. Richardson, et al. (2007). "Priority setting in health care using multi-attribute utility theory and programme budgeting and marginal analysis." *Social Science & Medicine* **64**(4).
- Phipps, S. (2002). "The well-being of young Canadian children in international perspective: A functionings approach." *REVIEW OF INCOME AND WEALTH*(4): 493-515.
- Protiere, C., C. Donaldson, et al. (2004). "The impact of information on non-health attributes on willingness to pay for multiple health care programmes." *Social Science & Medicine* **58**: 1257-1269.
- Qizilbash, M. (2002). "Development, Common Foes and Shared Values." *Review of Political Economy* **14**(4): 463-480.
- Qizilbash, M. (2002). On the Measurement of Human Development: Lecture prepared for the UNDP Training Course. UNDP, http://hdr.undp.org/docs/training/oxford/presentations/Qizilbash_HDIcritique.pdf.
- Raheim, M. and W. Haland (2006). "Lived Experience of Chronic Pain and Fibromyalgia: Women's Stories of Daily Life." *Qualitative Health Research* **16**(6): 741-761.
- Ratcliffe, J., K. J. Thomas, et al. (2006). "A randomised controlled trial of acupuncture care for persistent low back pain: cost effectiveness analysis." *BMJ: British Medical Journal* **333**(7569): 626.
- Rawls, J. (1999). *The Law of peoples*. Cambridge, Mass., Harvard University Press.

- Reinhardt, U. E. (1998). Abstracting from Distributional Effects, this Policy is Efficient. Health, Health Care and Health Economics: Prospectives on Distribution. M. L. Barer, T. E. Getzen and G. Stoddart. Chichester, Wiley.
- Richardson, J. (1994). "COST-UTILITY ANALYSIS - WHAT SHOULD BE MEASURED." SOCIAL SCIENCE & MEDICINE **39**(1): 7-21.
- Richardson, J. (2005). "Establishing the (extra)ordinary in chronic widespread pain." Health **9**(1): 31-48.
- Richardson, J., B. E. Ong, et al. (2006). "Is Chronic Widespread Pain Biographically Disruptive." Social Science & Medicine **63**: 1573-1585.
- Ritchie, J. and J. Lewis (2003). Qualitative Research Practice: A Guide for Social Science Students and Researchers. London, SAGE Publications.
- Rivero-Arias, O., H. Campbell, et al. (2005). "Surgical stabilisation of the spine compared with a programme of intensive rehabilitation for the management of patients with chronic low back pain: cost utility analysis based on a randomised controlled trial." BRITISH MEDICAL JOURNAL **330**(7502): 1239-1243.
- Rivero-Arias, O., A. Gray, et al. (2006). "Cost-utility analysis of physiotherapy treatment compared with physiotherapy advice in low back pain." SPINE **31**(12): 1381-1387.
- Robeyns, I. (2003). "Sen's capability approach and gender inequality: Selecting relevant capabilities." FEMINIST ECONOMICS **9**(2-3): 61-92.
- Robeyns, I. (2005). "The Capability Approach: a Theoretical Survey." Journal of Human Development **6**(1): 93-114.
- Robeyns, I. (2005b). "Selecting Capabilities for Quality of Life Measurement." Social Indicators Research **74**: 191-215.
- Robeyns, I. (2006). "The Capability Approach in Practice." The Journal of Political Philosophy **14**(3): 351-376.
- Ryan, M. and P. Shackley (1995). "Assessing the benefits of health care: how for should we go?" Quality in Health Care **4**: 207-213.
- Sach, T. H., R. Smith, D., et al. (2007). "A 'League Table' of Contingent Valuation Results for Pharmaceutical Interventions." PHARMACOECONOMICS **25**(2): 107-127.
- Sayar, K., M. Arian, et al. (2002). "Sleep Quality in Chronic Pain Patients." Canadian Journal of Psychiatry **47**(9): 844-848.
- Schischka, J., P. Dalziel, et al. (2008). "Applying Sen's Capability Approach to Poverty Alleviation Programs: Two Case Studies." Journal of Human Development **9**(2): 229-246.
- Schweikert, B., E. Jacobi, et al. (2006). "Effectiveness and cost-effectiveness of adding a cognitive behavioral treatment to the rehabilitation of chronic low back pain." JOURNAL OF RHEUMATOLOGY **33**(12): 2519-2526.
- Sen, A. (1983). "Poor, Relatively Speaking." Oxford Economic Papers **35**(2): 153-169.
- Sen, A. (1987). The Standard of Living. Cambridge, Cambridge University Press.
- Sen, A. (1992). Inequality Reexamined. Oxford, Clarendon Press.

- Sen, A. (1999). Commodities and Capabilities. New Delhi, Oxford University Press.
- Sen, A. (1999a). Development as Freedom. New York, Oxford University Press.
- Sen, A. (2002). "Health: Perception versus observation." BMJ **324**: 860-861.
- Sen, A. (2004). "Capabilities, lists, and public reason: Continuing the conversation." FEMINIST ECONOMICS **10**(3): 77-80.
- Shackley, P. and C. Donaldson (2002). "Should we use willingness to pay to elicit community preferences for health care? New evidence from using a 'marginal' approach." Journal of Health Economics **21**: 971-991.
- Sintonen, H. (2001). "The 15D Instrument of Health-Related Quality of Life: Properties and Applications." ANNALS OF MEDICINE **33**(5): 328-336.
- Smith, R., D. (2005). "Sensitivity to scale in contingent valuation: the importance of the budget constraint." Journal of Health Economics **24**: 515-529.
- Smith, R. and M. Dobson (1993). "Measuring Utility Values for QALYs: Two Methodological Issues." HEALTH ECONOMICS **2**: 349-355.
- Smith, R. D. and J. Richardson (2005a). "Can we estimate the 'social' value of a QALY?" Health Policy **74**(1): 77-84.
- Snelling, J. (1994). "The Effect of Chronic Pain on the Family Unit." Journal of Advanced Nursing **19**: 543-551.
- Sorensen, E. B. and S. Askegaard (2007). "Laddering: how (not) to do things with words." Qualitative Market Research: An International Journal **10**(1): 63-77.
- Sprangers, M. and C. E. Schwartz (1999). "Integrating response shift into health-related quality of life research: a theoretical model." Social Science & Medicine **48**: 1507-1515.
- Sugden, R. (1993). "Welfare, Resources, and Capabilities: A Review of Inequality Reexamined by Amartya Sen." Journal of Economic Literature **31**(4): 1947-1962.
- Sunstein, C. (2001). Cognition and Cost-Benefit Analysis. Cost Benefit Analysis: Legal, Economic, and Philosophical Perspectives. M. D. Adler and E. A. Posner. Chicago, University of Chicago Press.
- Svedsater, H. (2000). "Contingent Valuation of Global Environmental Resources: Test of perfect and regular embedding." Journal of Economic Psychology **21**: 605-623.
- Tarride, J. E., A. Gordon, et al. (2006). "Cost-effectiveness of pregabalin for the management of neuropathic pain associated with diabetic peripheral neuropathy and postherpetic neuralgia: a Canadian perspective." CLINICAL THERAPEUTICS **28**(11): 1922-1934.
- Taylor, R. J. and R. S. Taylor (2005). "Spinal cord stimulation for failed back surgery syndrome: A decision-analytic model and cost-effectiveness analysis." International Journal of Technology Assessment in Health Care **21**(3): 351-358.
- Thomas, E., G. Peat, et al. (2004). "The Prevalence of Pain and Pain Interference in a General Population of Older Adults: Cross-sectional Findings from the North Staffordshire Osteoarthritis Project (NorStop)." PAIN **110**: 361-368.

- Thomas, K. J., H. MacPherson, et al. (2005). "Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain." HEALTH TECHNOLOGY ASSESSMENT **9**(32): III-+.
- Thompson, M., M. Gawel, et al. (2005). "An economic evaluation of rizatriptan in the treatment of migraine." PHARMACOECONOMICS **23**(8): 837-850.
- Tolle, T., X. Xu, et al. (2006). "Painful diabetic neuropathy: a cross-sectional survey of health state impairment and treatment patterns." Journal of Diabetes and Its Complications **20**: 26-33.
- Torrance, G., D. Feeny, et al. (1996). "Multiattribute Utility Function for a Comprehensive Health Status Classification System: Health Utilities Index Mark 2." MEDICAL CARE **34**(7): 702-722.
- Torrance, G., W. Furlong, et al. (1995). "Multi-Attribute Preference Functions: Health Utilities Index." PHARMACOECONOMICS **7**(6): 503-520.
- Torrance, G. W. (1986). "Measurement of health state utilities for economic appraisal." J Health Econ **5**(1): 1-30.
- Tsuchiya, A. and A. Williams (2001). Welfare Economics and Economic Evaluation. Economic Evaluation in Health Care: Merging theory with practice. M. Drummond and A. McGuire. Oxford, Oxford University Press.
- Turner, J. A., M. Ersek, et al. (2005). "Self-Efficacy for managing Pain is Associated With Disability, Depression, and Pain Coping Among Retirement Community Residents with Chronic Pain." The Journal of Pain **6**(7): 471-479.
- Verkerk, M. A., J. J. V. Busschbach, et al. (2001). "Health-related quality of life research and the capability approach of Amartya Sen." QUALITY OF LIFE RESEARCH **10**(1): 49-55.
- Von Korf, M., J. Ormel, et al. (1992). "Grading the Severity of Chronic Pain." Pain **50**: 133-149.
- Werner, A., L. W. Isaksen, et al. (2004). "'I am not the kind of woman who complains of everything': Illness stories on self and shame in women with chronic pain." Social Science & Medicine **59**: 1035-1045.
- WHO (1946). Preamble to the Constitution of the World Health Organization. New York. **Official Records of the WHO, No. 2, p. 100.**
- Whynes, D. K., E. Frew, et al. (2003). "A Comparison of Two Methods for Eliciting Contingent Valuations of Colorectal Cancer Screening." Journal of Health Economics **22**: 555-574.
- Wigley, S. and A. Akkoyunlu-Wigley (2006). "Human capabilities versus human capital: Gauging the value of education in developing countries." SOCIAL INDICATORS RESEARCH **78**(2): 287-304.
- Willich, S. N., T. Reinhold, et al. (2006). "Cost-effectiveness of acupuncture treatment in patients with chronic neck pain." PAIN **125**(1-2): 107-113.
- Witt, C. M., S. Jena, et al. (2006). "Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain." AMERICAN JOURNAL OF EPIDEMIOLOGY **164**(5): 487-496.